

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Haploid Induction and Somatic Mutations of Potato

Permalink

<https://escholarship.org/uc/item/8nk358tm>

Author

Amundson, Kirk Richard

Publication Date

2021

Supplemental Material

<https://escholarship.org/uc/item/8nk358tm#supplemental>

Peer reviewed|Thesis/dissertation

HAPLOID INDUCTION AND SOMATIC MUTATIONS OF POTATO

By

KIRK RICHARD AMUNDSON
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PLANT BIOLOGY

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Luca Comai, Chair

Julin Maloof

Andrew Groover

Committee in Charge

2022

ABSTRACT

Haploid Induction and Somatic Mutations in Potato

By

Kirk Richard Amundson

Cultivated potato (*Solanum tuberosum* L.) is a highly heterozygous and clonally propagated autotetraploid ($2n=4x=48$) that exhibits severe inbreeding depression. These attributes of potato make improvement through conventional breeding difficult and slow, and have motivated development of alternative approaches. One attractive approach is to reinvent potato as a diploid, inbred line-based crop by capturing elite tetraploid germplasm at the diploid level via haploid induction. In potato, haploid induction is routinely achieved by pollination with specialized haploid inducer lines, which operate by an undetermined mechanism. Genome editing represents another attractive approach. Currently, editing requires regeneration of individually edited cells into plants, which has long been known to destabilize the genome. An improved understanding of the biological processes underlying haploid induction and genome instability could aid in techniques to improve potato that may be broadly applicable in plants. Studies were conducted to better understand the frequency, extent and underlying basis of genome instability in potato haploid induction or tissue culture regeneration. Sequencing the genomes of 1,086 primary dihaploids revealed whole-chromosome aneuploidy in 8% of progeny. In the majority of cases, aneuploidy was a single additional chromosome from the tetraploid maternal parent, likely a result of meiotic nondisjunction. Chromosomes from the haploid inducer parent were detected in 0.5% of progeny and showed evidence of restructuring. Among progeny with additional chromosomes from the haploid inducer, additional inducer-derived DNA segments were detected, but their location could not be precisely determined, and in some cases, were artifacts

of reference genome assembly. Genome sequencing of 134 triploid or tetraploid hybrids obtained from potato haploid induction crosses revealed ploidy-dependent genome instability of the haploid inducer parent: inducer chromosomes were stable in triploids, but not in tetraploids. Tetraploid hybrids could be produced by several possible mechanisms, but most were produced by first meiotic division restitution of the haploid inducer. This study revealed that fertilization can occur in potato haploid induction, ruling out parthenogenesis as an exclusive mechanism; however, the vast majority of primary dihaploids were free of detectable haploid inducer DNA. In light of these findings, the incidental transfer of haploid inducer DNA to primary dihaploids, which was once thought to be both pervasive and undesirable for potato breeding, occurs infrequently. Analysis of 12 leaf protoplast regenerants from a single tetraploid cultivar revealed a preexisting unbalanced translocation, tr8-7, in the protoplast donor. Genetic and cytogenetic analyses indicated that tr8-7 is the derived state within the protoplast donor, and that cells carrying tr8-7 compose the L2 and L3 cell layers of the shoot apical meristem. Regeneration also led to whole-chromosome aneuploidy, copy-neutral change of heterozygosity consistent with chromosome substitution, as well as catastrophic chromosome shattering similar to that observed in cancer genomes. These findings provide a framework for studying somatic mutations in long-lived and/or polyploid plants and provide further evidence of the destabilizing effect of tissue culture regeneration.

Dedicated to Mom, whose casual suggestion to pursue research left a lasting impact.

ACKNOWLEDGEMENTS

I thank Drs. Luca Comai and Isabelle Henry for their mentorship during my research.

I thank the members of my committee, Drs. Luca Comai, Julin Maloof and Andrew Groover for their expertise, guidance and direction.

I thank my collaborators Monica Santayana, Drs. Ek Han Tan, Meredith Bonierbale, Awais Khan, Hannele Lindqvist-Kreuze, Jiming Jiang, Guilherme Tomaz Braz, and Xin Zhao for their contributions to this research, and members of the Comai laboratory, especially Benny Ordoñez, for lending their expertise and support.

I thank all of the interns that have helped me push this research forward: Sophia Jinata, Dana Ngo, Jonathan Do, Lisa Malins, Isabelle DeMarco and Cleo Yang.

I thank my mentors during my undergraduate education: Drs. M. David Marks, Kevin Dorn and Deena Wassenberg, for taking a chance on an eager undergraduate without any research or teaching experience. I wouldn't be here without your training and encouragement.

I thank my family for their support of my education, and more importantly, my wellbeing and happiness.

I thank the members of the Plant Biology Graduate Group for their friendship and support.

I thank my girlfriend Sarah Guest for her patience, support, constructive criticism and reminders to take breaks.

TABLE OF CONTENTS

TABLE OF CONTENTS	vi
List of Tables	x
List of Figures	xi
Chapter 1	1
Introduction	1
Challenges of potato breeding	2
New technologies to improve potato breeding: Diploid breeding	2
Bottleneck to diploid breeding: Haploid Induction	3
New technologies to improve potato breeding: Genetic engineering	6
Bottleneck to genetic engineering: Regeneration and Genome Instability	7
Problem Definition	8
Objectives	9
Dissertation Outline	9
Literature Cited	10
Chapter 2	17
Genomic Outcomes of Haploid Induction Crosses in Potato (<i>Solanum tuberosum</i> L.)	17
Abstract	18
Summary	18
Introduction	19
Results	23
Induction, selection and sequencing of dihaploids	23
Maternally derived trisomy	23
Determining which SNP bins are informative	25
CNV analysis	26
SNP loci consistent with addition or introgression were rare and dispersed	27
Discussion	29
Conclusions	33
Materials and Methods	33
Plant material	33
Genomic DNA library preparation, sequencing, and pre-processing	34
Variant calling	34
Dosage analysis	35
Cytological analysis	37
Data availability	38

Funding	38
Acknowledgements	38
Author Contributions	38
Literature Cited	39
Figures	44
Supplemental material	50
Supplemental Figures	50
Supplemental Tables	65
Supplemental Data Sets	65
Chapter 3	66
Rare instances of haploid inducer DNA in potato dihaploids and ploidy-dependent genome instability	66
Abstract	67
Introduction	67
Results	70
Widespread aneuploidy among primary dihaploids	70
Retention of haploid inducer chromosomes	72
Detection of haploid inducer-derived DNA segments in dihaploids	73
Selective instability of the haploid inducer genome in dihaploids and tetraploid hybrids	74
Tetraploid hybrids produced by first meiotic division restitution of the haploid inducer	75
Discussion	76
Materials and Methods	81
Plant material	81
Flow cytometry	82
Whole genome resequencing	82
Variant calling	83
Chromosome dosage analysis	84
High resolution analysis of parental DNA contribution	85
Low-pass haplotype analysis	85
Statistical analyses	86
Proportion of aneuploids among dihaploids by female parent	86
Fisher exact counts of HI dihaploid introgression events by haploid inducer	86
Linkage disequilibrium from dosage variable states	86
Logistic regression model for incidence of paternal (maternal) genome instability in triploid and tetraploid hybrids	87
Data Availability	88
Acknowledgements	88
Author Contributions	88

Conflict of Interest Statement	89
Literature Cited	89
Figures	94
Tables	100
Supplemental Material	103
Supplemental Figures	103
Supplemental Tables	112
Supplemental Data Sets	114
Chapter 4	115
Genetic mechanisms underlying somatic chromosomal changes in potato	115
Abstract	115
Introduction	116
Results	120
PI310467 displayed an unbalanced chromosome translocation	120
Somatic loss of unbalanced translocation observed among regenerants	121
Mosaicism of PI310467 resulted in two genetically distinct regenerant types	122
Chromoanagenesis	125
Chromosome substitution	127
Discussion	128
Chromosome substitution	129
Chromoanagenesis	130
PI310467 is not Desiree	131
Unexpected mosaicism of PI310467	131
Materials and Methods	134
Plant material	134
Flow cytometry	135
Library construction and sequencing	135
Variant calling	136
Chromosome dosage	137
Phylogenetic analysis	137
FISH	139
Simulations	140
Data Availability	140
Acknowledgements	141
Author Contributions	141
Conflict of Interest Statement	141
Literature Cited	141

Figures	150
Supplemental Material	162
Supplemental Figures	162
Supplemental Tables	173
Supplemental Data Sets	173
Chapter 5	174
General Conclusions	174
Overview of Dissertation Research	175
Literature Cited	179

List of Tables

Supplemental Table S2.1 Summary of sequencing data.	65
Supplemental Table S2.2 Rare dosage variants and associated haploid inducer SNP profiles	65
Table 3.1 Putative introgressions of haploid inducer DNA segments in dihaploid potatoes.	100
Table 3.2 Frequency of chromosome breakage among progeny of potato haploid induction crosses.	102
Supplemental Table S3.1 Ploidy and embryo spot phenotype of progeny.	112
Supplemental Table S3.2 Description of 19 potato clones used for dihaploid extraction.	113
Supplemental Table S4.1 Counts of PI310467 genotypes at Desiree-homozygous loci.	173

List of Figures

Figure 2.1 Production of the LOP population and expected types	44
Figure 2.2 Aneuploidy detected in 167 primary dihaploids by genome sequencing	45
Figure 2.3 Chromosomal distributions of dosage variation in the LOP dihaploid population	47
Figure 2.4 Example of segregating CNV among the LOP dihaploid population	49
Supplemental Figure S2.1 Aneuploidy detection	50
Supplemental Figure S2.2 Power analysis for HI SNP measured in dihaploids by low pass sequencing	52
Supplemental Figure S2.3 Expanded view of dosage variation in dihaploid population	53
Supplemental Figure S2.4 Assignment of CNV type to each dihaploid line	58
Supplemental Figure S2.5 CNV inference from dosage-variable states	59
Supplemental Figure S2.6 Power analysis for parental origin of rare structural variants	60
Supplemental Figure S2.7 Haploid inducer allele support at putative introgression loci	61
Supplemental Figure S2.8 Comparison of putative introgression sites in three dihaploids	62
Supplemental Figure S2.9 Marker conversion rate putative introgression events	63
Supplemental Figure S2.10 Screenshot of representative false positive SNP locus	64
Figure 3.1 Possible outcomes of potato haploid induction crosses	94
Figure 3.2 Incidence and parental origin of aneuploidy among putative dihaploids	95
Figure 3.3 Paternal genomic contribution to maternal dihaploids	96
Figure 3.4 Haploid inducer genome instability in tetraploid potato hybrids	97
Figure 3.5 Tetraploid hybrid formation in potato haploid induction	99

Supplemental Figure S3.1 Chromosome affected by state of loss or gain of potato dihaploids	103
Supplemental Figure S3.2 Aneuploidy frequency partitioned by maternal parent	104
Supplemental Figure S3.3 Aneuploidy frequency partitioned by paternal parent	105
Supplemental Figure S3.4 Inferred karyotypes of eight haploid inducer addition dihaploids	106
Supplemental Figure S3.5 Haploid inducer allele representation in dihaploids	107
Supplemental Figure S3.6 Dosage variation linkage matrix of WA.077 dihaploids	108
Supplemental Figure S3.7 Test of potato haploid inducer (HI) haplotype phasing	109
Supplemental Figure S3.8 Centromeric heterozygosity of tetraploid hybrids	110
Supplemental Figure S3.9 Inferred karyotype of disomic chromosome 11 tetraploid hybrid	111
Figure 4.1 Somatic loss of translocation among protoplast regenerants of PI310467	150
Figure 4.2 Recurring tr8-7 loss due to PI310467 mosaicism	152
Figure 4.3 SNP alleles associated with tr8-7 loss are common among potatoes	154
Figure 4.4 Chromosomal rearrangements of regenerant MF74	156
Figure 4.5 Copy-neutral change of heterozygosity consistent with chromosome substitution	158
Figure 4.6 Model of PI310467 chromosomal rearrangement and developmental events	160
Supplemental Figure S4.1 Classification of dihaploids, aneuploids, selfs and hybrids	162
Supplemental Figure S4.2 Dihaploid linkage analysis	163
Supplemental Figure S4.3 PI310467 S1 population linkage analysis	164
Supplemental Figure S4.4 Dosage plots of three independent cuttings of PI310467	165
Supplemental Figure S4.5 Read depth analysis of regenerated clones	166

Supplemental Figure S4.6 Comparison of panel clones with multiple entries	167
Supplemental Figure S4.7 PI310467 heterozygosity at Desiree-homozygous SNP loci	168
Supplemental Figure S4.8 Expanded view of PI310467 and MF74 read depth	169
Supplemental Figure S4.9 Change of heterozygosity from simulated haplotype substitution	171

Chapter 1

Introduction

Challenges of potato breeding

Potato (*Solanum tuberosum* L.) is the third most important food crop in terms of direct human consumption after wheat and rice, and is the most important vegetable crop (Birch et al., 2012). While potatoes are primarily grown for processed foods in the United States (Lin et al., 2001), potato cultivation has expanded substantially to become a cornerstone of food security in the developing world, notably in Asia and sub-Saharan Africa (Birch et al., 2012). An acceptable cultivar balances producer and consumer demands for high yield, tuber quality, nutritional content and resilience to current and emerging biological stresses, forcing potato breeders to select many traits at once (Bonierbale et al., 2020). Outside of equatorial South America, cultivated potatoes are highly heterozygous autotetraploids ($2n=4x=48$) that show acute inbreeding depression when selfed (Zhang et al., 2019; De Jong and Rowe, 1971). These aspects of the potato genome have limited the scope of conventional potato breeding. Typically, heterozygous tetraploid parents are crossed, and 100,000 or more first-generation offspring are grown clonally and phenotypically selected over the course of a decade or more. With limited opportunity for recombination or the ability to inbreed, potato's enhancement lags behind that of other crops, and yield improvement due to breeding has been stagnant over the last century (Douches et al., 1996). The many traits demanded of elite potato cultivars, coupled with the challenges posed by their complex genomes have motivated the development of alternative improvement strategies.

New technologies to improve potato breeding: Diploid breeding

An emerging strategy involves the reinvention of potato as a diploid inbred line-based crop (Lindhout et al., 2011; Jansky et al., 2016; Jansky and Spooner, 2017; Chase, 1963). Converting

potato to a diploid crop requires developments on many fronts, such as overcoming S-RNase-mediated self-incompatibility (Clot et al., 2020; Enciso-Rodriguez et al., 2019; Ye et al., 2018), purging a high genetic load (Zhang et al., 2021, 2019), and reproduction through botanical seeds instead of tubers. A diploid potato could have multiple benefits. Compared to tetraploid breeding, diploid breeding would shorten the breeding cycle, simplify genetic mapping, enable breeding for recessive and/or complex traits, ease access to useful traits from the richly diverse wild relatives of potato, avoid pathogen buildup from continuous vegetative propagation, and enable storage and shipment of botanical seeds in place of “seed tubers” or in vitro cuttings (Jansky et al., 2016). The proposed first step is to capture the genetic diversity of elite tetraploid germplasm at the diploid level.

Bottleneck to diploid breeding: Haploid Induction

In potato, ploidy reduction is routinely achieved by pollination of a tetraploid variety of interest with specialized diploid ($2n=2x=24$) varieties that act as haploid inducers (Hermsen and Verdenius, 1973; Hutten et al., 1993; Ordoñez et al., 2021). Some of the resulting progeny are $2n=2x=24$ and, by convention, are called primary dihaploids to indicate both a $2x$ chromosome number and immediate descent from a tetraploid parent. The mechanism by which primary dihaploids arise in potato haploid induction crosses remains unclear, but is relevant in light of recent efforts on diploid breeding.

An important question is whether double fertilization occurs or not. In a $4x \times 2x$ cross, double fertilization is expected to produce a triploid ($2n=3x=36$) zygote and $5x (=60)$ primary endosperm. If the $2x$ parent instead contributes an unreduced gamete, which carries two $2x$ sperm, the expected ploidy of the zygote and primary endosperm are instead $4x$ and $6x$,

respectively. Cytological analysis revealed that most developing seeds produced by the 4x by 2x (HI) cross of potato do not survive beyond 2-3 weeks after pollination, which agreed with later reports of a strong triploid block in 4x by 2x crosses of potato (Marks, 1966; Jackson et al., 1978). Furthermore, nearly all (91/92) surviving ovules contained hexaploid (6x) endosperm associated with either embryos that were either 2x or 4x, or with no embryo at all (Wangenheim et al., 1960). To explain the association of 2x embryos with 6x endosperm, Wangenheim et al. proposed that two 1x sperm (or their equivalent) fertilized the central cell and the viable endosperm stimulated parthenogenesis of the unfertilized egg. The induced parthenogenesis hypothesis was accepted by potato researchers (Hermsen and Verdenius, 1973) and appears in more recent reviews (Jansky and Spooner, 2017). However, investigations of dihaploid potatoes suggested that haploid inducer DNA can appear in dihaploid or near-dihaploid aneuploids (Clulow et al., 1991; Waugh et al., 1992; Clulow et al., 1993; Wilkinson et al., 1995; Allainguillaume et al., 1997; Clulow and Rousselle-Bourgeois, 1997; Straadt and Rasmussen, 2003; Ercolano et al., 2004; Pham et al., 2019), prompting an alternative hypothesis: egg fertilization followed by occasionally incomplete elimination of the pollinator genome. In certain wide crosses, *Hordeum vulgare* x *H. bulbosum*, for example, or in interspecific crosses where one parent contains alterations to CENTROMERIC HISTONE H3 (CENH3), hybrid zygotes show laggard chromosomes at mitotic anaphase, micronuclei, and loss of some or all chromosomes of the haploid inducer parent during embryogenesis (Ravi and Chan, 2010; Sanei et al., 2011; Tan et al., 2015; Laurie and Bennett, 1986; Ishii et al., 2016 and references therein). Haploid induction is often incompletely penetrant, meaning that diploid and near-diploid aneuploid progeny are also obtained. In near-diploid aneuploids produced by CENH3-mediated haploid induction, only the chromosomes inherited from the haploid inducer parent are lost,

duplicated or restructured (Tan et al., 2015; Kuppu et al., 2015; Maheshwari et al., 2015). In the potato haploid induction cross, triploid and tetraploid hybrids are also represented among the progeny, but whether selective instability of inducer-derived chromosomes is not known. This poses a second question that could potentially shed light on the mechanism of haploid induction in potato: what is the fate of HI-derived chromosomes in hybrids?

A deeper understanding of haploid induction in potato has important implications for breeding potato and other crops. First, the potato haploid inducers are poorly adapted for global cultivation because they exhibit traits such as short tuber dormancy, short-day tuberization and small, irregular tubers. Evidence from previous studies indicates that incidental transfer of HI DNA can occur, and genes from the retained inducer DNA can be expressed (Clulow et al., 1993) and impact plant phenotype (Allainguillaume et al., 1997). As these studies are limited to relatively small cohorts, the frequency of haploid inducer DNA retention, and the structure of the retained DNA remain open. These questions are especially pertinent as the public and private sector have begun inducing dihaploids of elite tetraploid cultivars (Jansky et al., 2016). The second, broader, implication is that naturally occurring haploid inducers are available for relatively few crop species. Haploid inducers enable rapid generation of true-breeding material for genetic mapping (Seymour et al., 2012) and inbred line development, providing a powerful resource in the few crop species that have them. Recent identification of the genes underlying haploid induction in maize (Kelliher et al., 2017; Gilles et al., 2017; Liu et al., 2017; Zhong et al., 2019) has resulted in engineered haploid inducers of rice (Yao et al., 2018). The major effect gene, MATRILINEAL/NOT LIKE DAD, is functionally conserved among grasses (Yao et al., 2018), but lacks a clear eudicot ortholog. While the minor effect gene, ZmDMP, is functionally

conserved in eudicots (Wang et al., 2021; Zhong et al., 2021, 2020), its effect on haploid induction is relatively modest. Elucidating the molecular players of haploid induction in potato may provide an avenue toward efficient haploid breeding in more crop species.

New technologies to improve potato breeding: Genetic engineering

Ushered in by the discovery of the bacterial immunity system CRISPR-Cas9 as a customizable site-specific nuclease (Jinek et al., 2012) and its development into an *in vivo* gene editing tool, the past few years have witnessed an explosion in the application of genome editing across a broad range of plant species. CRISPR reagents are delivered into plant cells as DNA, RNA or protein-RNA that assemble into active site-directed nucleases to generate targeted double-stranded breaks. Plant cells can repair these DSBs by non-homologous end joining (NHEJ) or by integrating a different piece of DNA at the DSB through homology-dependent repair pathways. The former repair pathway efficiently generates knockout mutations, as indels of a few base pairs can result from error-prone repair. In contrast, homology-based repair enables more sophisticated modifications. Efficient genome editing techniques are already available in potato, and have been deployed for modifying tuber starch content (Johansen et al., 2019; Andersson et al., 2017), conferring late blight resistance (Kieu et al., 2021), engineering self compatibility (Enciso-Rodriguez et al., 2019; Ye et al., 2018), and improving cold storage tolerance (Clasen et al., 2016). Even as genome editing technology matures, conventional transgenesis continues to see use in important applications, such as engineered durable late blight resistance (Ghislain et al., 2019).

Bottleneck to genetic engineering: Regeneration and Genome Instability

Regardless of the type of editing reagent and the change made, editing takes place in individual plant cells. Tissue culture regeneration is then required to recover a transgenic or gene-edited plant. Despite over 50 years of technological advances in plant tissue culture, regeneration remains a severe bottleneck that hinders advances in genome editing reagents, reagent delivery into plant cells, and identification of loci to edit. This bottleneck exists for several reasons. For example, tissue culture conditions often need to be optimized empirically for different species, as well as genotypes within species. Regeneration is both technically demanding and labor-intensive, and as a result, transformation of several important crop species, maize for example, is carried out in a handful of specialized laboratories (Altpeter et al., 2016). Furthermore, callus regeneration is mutagenic in virtually all plant species studied to date (Neelakandan and Wang, 2012; Veilleux and Johnson, 2010; Phillips et al., 1994; Lee and Phillips, 1988; Larkin and Scowcroft, 1981). There are numerous reports of changes to DNA sequence (Park et al., 2020; Tang et al., 2018; Zhang et al., 2014; Miyao et al., 2012; Jiang et al., 2011), DNA methylation (Han et al., 2018; Stelpflug et al., 2014; Stroud et al., 2013; Vining et al., 2013; Kaeppler and Phillips, 1993a, 1993b), transposable elements (Miyao et al., 2012; Sato et al., 2011; Kikuchi et al., 2003; Hirochika et al., 1996; Peschke and Phillips, 1991; Brettell and Dennis, 1991; Peschke et al., 1987), chromosome number and chromosome structure (Karp et al., 1987; Wheeler et al., 1985; Ramulu et al., 1983; Karp et al., 1982; Pucker et al., 2019; Fossi et al., 2019; Gernand et al., 2007; Gill et al., 1987; Lee and Phillips, 1988) among tissue culture regenerated plants. While much has been recorded on tissue culture induced variation as a phenomenon, mechanistic insight into how genome instability occurs in tissue culture remains limited.

Potato is an attractive system for studying somaclonal variation due to its autopolyploid genome, relative ease of regeneration from protoplasts or various explant types, and vegetative propagation as tubers or *in vitro* cuttings. Potatoes are vegetatively propagated for extended periods of time, and many spontaneous bud sports, some of which arose long ago, are still widely grown today (Bethke et al., 2014; Miller et al., 1999; Miller, 1954). For example, Russet Burbank, a sport of a variety released in 1876, is still the most widely grown cultivar in the United States, and the underlying mutation is not known (Bethke et al., 2014). Deeper study of the genomic changes incurred during regeneration could provide insight into the mechanistic basis of somatic chromosomal change both inside and outside of tissue culture. Ample genomic and cytogenetic resources are available: chromosome-specific FISH probe sets (Braz et al., 2018; He et al., 2018), multiple long-read genome assemblies (Sun et al., 2021; Pham et al., 2020; Zhou et al., 2020; van Lieshout et al., 2020), haploid inducers (Ordoñez et al., 2021; Hermsen and Verdenius, 1973; Hutten et al., 1993), and a wealth of genotype and sequencing data (Prodhomme et al., 2020; Sharma et al., 2018; Pham et al., 2017; Hardigan et al., 2017; Hirsch et al., 2013). These tools enable further study into somatic mutations, whether artificially induced or naturally occurring, and inference and comparison of their underlying mechanisms.

Problem Definition

Consumer, producer and environmental pressures make tetraploid potato breeding increasingly challenging, and motivate alternative strategies for its ongoing improvement. One proposed strategy is to reinvent potatoes as a diploid inbred crop (Jansky et al. 2016). A critical first step of this approach is to capture genetic diversity of elite tetraploid cultivars at the diploid level, which is routinely achieved by pollination with haploid inducers. Although haploid induction

was originally proposed to occur by parthenogenesis, later detection of inducer-specific genetic markers in haploid and near-haploid aneuploids suggest postzygotic chromosome elimination as an alternate mechanism (Clulow et al. 1991). The extent and frequency of incidental haploid inducer retention, and the structure of the retained inducer DNA, were not known. Another common strategy is improvement through transgenesis or gene editing, which requires transformed or edited cells to be regenerated into whole plants. Although it is well established that this regeneration changes the genome, the mechanistic basis of these changes is unclear. An improved understanding of the biological mechanisms underlying genome instability in haploid induction and regeneration may aid in the understanding of plant reproduction and mechanisms driving genomic change, and in efforts to innovate potato breeding.

Objectives

1. Determine the extent and parental origin of aneuploidy observed in dihaploid induction crosses of potato.
2. Investigate the frequencies of chromosome instability in hybrid byproducts of the potato dihaploid induction cross.
3. Investigate the spectrum of novel karyotypes obtained from potato clones regenerated from leaf protoplasts.

Dissertation Outline

Chapter 1 is an introduction to haploid induction, genome instability and potato genome biology that outlines the scope of the questions addressed in this dissertation.

Chapter 2 is a sequencing-based survey of aneuploidy and genomic variation of 167 dihaploids produced by pollination of a single tetraploid landrace clone with haploid inducer clones.

Chapter 3 is a sequencing-based survey of genomic variation of 919 dihaploids, 30 triploid hybrids, and 134 tetraploid hybrids produced from pollination of various tetraploid breeding lines with haploid inducer varieties.

Chapter 4 is a study using deep whole genome sequencing and cytological analyses to define the types and extent of genome instability observed among plants regenerated from single leaf protoplasts in tissue culture. Chromosomal variation that existed among cells of the protoplast donor or was induced by regeneration were documented.

Chapter 5 summarizes the findings of this work.

Literature Cited

- Allainguillaume, J., Wilkinson, M.J., Clulow, S.A., and Barr, S.N.R.** (1997). Evidence that genes from the male parent may influence the morphology of potato dihaploids. *Theor. Appl. Genet.* **94**: 241–248.
- Altpeter, F. et al.** (2016). Advancing Crop Transformation in the Era of Genome Editing. *Plant Cell* **28**: 1510–1520.
- Andersson, M., Turesson, H., Nicolia, A., Fält, A.-S., Samuelsson, M., and Hofvander, P.** (2017). Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Rep.* **36**: 117–128.
- Bethke, P.C., Nassar, A.M.K., Kubow, S., Leclerc, Y.N., Li, X.-Q., Haroon, M., Molen, T., Bamberg, J., Martin, M., and Donnelly, D.J.** (2014). History and Origin of Russet Burbank (Netted Gem) a Sport of Burbank. *American Journal of Potato Research* **91**: 594–609.
- Birch, P.R.J., Bryan, G., Fenton, B., Gilroy, E.M., Hein, I., Jones, J.T., Prashar, A., Taylor, M.A., Torrance, L., and Toth, I.K.** (2012). Crops that feed the world 8: Potato: are the trends of increased global production sustainable? *Food Secur.* **4**: 477–508.

- Bonierbale, M.W., Amoros, W.R., Salas, E., and de Jong, W.** (2020). Potato Breeding. *The Potato Crop*: 163–217.
- Braz, G.T., He, L., Zhao, H., Zhang, T., Semrau, K., Rouillard, J.-M., Torres, G.A., and Jiang, J.** (2018). Comparative Oligo-FISH Mapping: An Efficient and Powerful Methodology To Reveal Karyotypic and Chromosomal Evolution. *Genetics* **208**: 513–523.
- Brettell, R.I.S. and Dennis, E.S.** (1991). Reactivation of a silent Ac following tissue culture is associated with heritable alterations in its methylation pattern. *Molecular and General Genetics MGG* **229**: 365–372.
- Chase, S.S.** (1963). ANALYTIC BREEDING IN SOLANUM TUBEROSUM L. – A SCHEME UTILIZING PARTHENOTES AND OTHER DIPLOID STOCKS. *Can. J. Genet. Cytol.* **5**: 359–363.
- Clasen, B.M. et al.** (2016). Improving cold storage and processing traits in potato through targeted gene knockout. *Plant Biotechnol. J.* **14**: 169–176.
- Clot, C.R., Polzer, C., Prodhomme, C., Schuit, C., Engelen, C.J.M., Hutten, R.C.B., and van Eck, H.J.** (2020). The origin and widespread occurrence of Sli-based self-compatibility in potato. *Theor. Appl. Genet.* **133**: 2713–2728.
- Clulow, S.A. and Rousselle-Bourgeois, F.** (1997). Widespread introgression of Solanum phureja DNA in potato (*S. tuberosum*) dihaploids. *Plant Breed.* **116**: 347–351.
- Clulow, S.A., Wilkinson, M.J., and Burch, L.R.** (1993). Solanum phureja genes are expressed in the leaves and tubers of aneusomatic potato dihaploids. *Euphytica* **69**: 1–6.
- Clulow, S.A., Wilkinson, M.J., Waugh, R., Baird, E., Demaine, M.J., and Powell, W.** (1991). Cytological and molecular observations on Solanum phureja-induced dihaploid potatoes. *Theor. Appl. Genet.* **82**: 545–551.
- De Jong, H. and Rowe, P.R.** (1971). Inbreeding in cultivated diploid potatoes. *Potato Res.* **14**: 74–83.
- Douches, D.S., Maas, D., Jastrzebski, K., and Chase, R.W.** (1996). Assessment of potato breeding progress in the USA over the last century. *Crop Sci.* **36**: 1544–1552.
- Enciso-Rodriguez, F., Manrique-Carpintero, N.C., Nadakuduti, S.S., Buell, C.R., Zarka, D., and Douches, D.** (2019). Overcoming Self-Incompatibility in Diploid Potato Using CRISPR-Cas9. *Front. Plant Sci.* **10**: 376.
- Ercolano, M.R., Carputo, D., Li, J., Monti, L., Barone, A., and Frusciante, L.** (2004). Assessment of genetic variability of haploids extracted from tetraploid ($2n = 4x = 48$) *Solanum tuberosum*. *Genome* **47**: 633–638.
- Fossi, M., Amundson, K.R., Kuppu, S., Britt, A.B., and Comai, L.** (2019). Regeneration of *Solanum tuberosum* plants from protoplasts induces widespread genome instability. *Plant Physiol.*
- Gernand, D., Golczyk, H., Rutten, T., Ilnicki, T., Houben, A., and Joachimiak, A.J.** (2007). Tissue culture triggers chromosome alterations, amplification, and transposition of repeat sequences in *Allium fistulosum*. *Genome* **50**: 435–442.
- Ghislain, M., Byarugaba, A.A., Magembe, E., Njoroge, A., Rivera, C., Román, M.L., Tovar, J.C.,**

- Gamboa, S., Forbes, G.A., Kreuze, J.F., Barekye, A., and Kiggundu, A.** (2019). Stacking three late blight resistance genes from wild species directly into African highland potato varieties confers complete field resistance to local blight races. *Plant Biotechnol. J.* **17**: 1119–1129.
- Gill, B.S., Kam-Morgan, L.N.W., and Shepard, J.F.** (1987). Cytogenetic and phenotypic variation in mesophyll cell-derived tetraploid potatoes. *J. Hered.* **78**: 15–20.
- Gilles, L.M. et al.** (2017). Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. *EMBO J.*: e201796603.
- Han, Z., Crisp, P.A., Stelpflug, S., Kaeppler, S.M., Li, Q., and Springer, N.M.** (2018). Heritable Epigenomic Changes to the Maize Methylome Resulting from Tissue Culture. *Genetics* **209**: 983–995.
- Hardigan, M.A., Laimbeer, F.P.E., Newton, L., Crisovan, E., Hamilton, J.P., Vaillancourt, B., Wiegert-Rininger, K., Wood, J.C., Douches, D.S., Farré, E.M., Veilleux, R.E., and Buell, C.R.** (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. U. S. A.* **114**: E9999–E10008.
- He, L., Braz, G.T., Torres, G.A., and Jiang, J.** (2018). Chromosome painting in meiosis reveals pairing of specific chromosomes in polyploid *Solanum* species. *Chromosoma* **127**: 505–513.
- Hermesen, J.G.T. and Verdenius, J.** (1973). Selection from *Solanum tuberosum* group Phureja of genotypes combining high-frequency haploid induction with homozygosity for embryo-spot. *Euphytica* **22**: 244–259.
- Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H., and Kanda, M.** (1996). Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 7783–7788.
- Hirsch, C.N., Hirsch, C.D., Felcher, K., Coombs, J., Zarka, D., Van Deynze, A., De Jong, W., Veilleux, R.E., Jansky, S., Bethke, P., Douches, D.S., and Buell, C.R.** (2013). Retrospective view of North American potato (*Solanum tuberosum* L.) breeding in the 20th and 21st centuries. *G3* **3**: 1003–1013.
- Hutten, R.C.B., Scholberg, E.J.M.M., Huigen, D.J., Hermesen, J.G.T., and Jacobsen, E.** (1993). Analysis of dihaploid induction and production ability and seed parent x pollinator interaction in potato. *Euphytica* **72**: 61–64.
- Ishii, T., Karimi-Ashtiyani, R., and Houben, A.** (2016). Haploidization via Chromosome Elimination: Means and Mechanisms. *Annu. Rev. Plant Biol.* **67**: 421–438.
- Jackson, M.T., Rowe, P.R., and Hawkes, J.G.** (1978). Crossability relationships of andean potato varieties of three ploidy levels. *Euphytica* **27**: 541–551.
- Jansky, S.H. et al.** (2016). Reinventing Potato as a Diploid Inbred Line–Based Crop. *Crop Sci.* **56**: 1412–1422.
- Jansky, S.H. and Spooner, D.M.** (2017). The evolution of potato breeding. *Plant Breed. Rev.*
- Jiang, C., Mithani, A., Gan, X., Belfield, E.J., Klingler, J.P., Zhu, J.-K., Ragoussis, J., Mott, R., and Harberd, N.P.** (2011). Regenerant *Arabidopsis* lineages display a distinct genome-wide spectrum of mutations conferring variant phenotypes. *Curr. Biol.* **21**: 1385–1390.

- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E.** (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*.
- Johansen, I.E., Liu, Y., Jørgensen, B., Bennett, E.P., Andreasson, E., Nielsen, K.L., Blennow, A., and Petersen, B.L.** (2019). High efficacy full allelic CRISPR/Cas9 gene editing in tetraploid potato. *Scientific Reports* **9**.
- Kaeppler, S.M. and Phillips, R.L.** (1993a). DNA methylation and tissue culture-induced variation in plants. *In Vitro Cellular & Developmental Biology - Plant* **29**: 125–130.
- Kaeppler, S.M. and Phillips, R.L.** (1993b). Tissue culture-induced DNA methylation variation in maize. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 8773–8776.
- Karp, A., Jones, M.G.K., Ooms, G., and Bright, S.W.J.** (1987). Potato Protoplasts and Tissue Culture in Crop Improvement. *Biotechnol. Genet. Eng. Rev.* **5**: 1–32.
- Karp, A., Nelson, R.S., Thomas, E., and Bright, S.W.** (1982). Chromosome variation in protoplast-derived potato plants. *Theor. Appl. Genet.* **63**: 265–272.
- Kelliher, T. et al.** (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* **542**: 105–109.
- Kieu, N.P., Lenman, M., Wang, E.S., Petersen, B.L., and Andreasson, E.** (2021). Mutations introduced in susceptibility genes through CRISPR/Cas9 genome editing confer increased late blight resistance in potatoes. *Scientific Reports* **11**.
- Kikuchi, K., Terauchi, K., Wada, M., and Hirano, H.-Y.** (2003). The plant MITE mPing is mobilized in anther culture. *Nature* **421**: 167–170.
- Kuppu, S., Tan, E.H., Nguyen, H., Rodgers, A., Comai, L., Chan, S.W.L., and Britt, A.B.** (2015). Point Mutations in Centromeric Histone Induce Post-zygotic Incompatibility and Uniparental Inheritance. *PLoS Genet.* **11**: e1005494.
- Larkin, P.J. and Scowcroft, W.R.** (1981). Somaclonal variation - a novel source of variability from cell cultures for plant improvement. *Züchter Genet. Breed. Res.* **60**: 197–214.
- Laurie, D.A. and Bennett, M.D.** (1986). Wheat × maize hybridization. *Can. J. Genet. Cytol.* **28**: 313–316.
- Lee, M. and Phillips, R.L.** (1988). The chromosomal basis of somaclonal variation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **39**: 413–437.
- van Lieshout, N., van der Burgt, A., de Vries, M.E., Ter Maat, M., Eickholt, D., Esselink, D., van Kaauwen, M.P.W., Kodde, L.P., Visser, R.G.F., Lindhout, P., and Finkers, R.** (2020). Solyntus, the New Highly Contiguous Reference Genome for Potato (*Solanum tuberosum*). *G3* **10**: 3489–3495.
- Lin, B.-H., Lucier, G., Allshouse, J., and Kantor, L.S.** (2001). Market Distribution of Potato Products in the United States. *J. Food Prod. Mark.* **6**: 63–78.
- Lindhout, P., Meijer, D., Schotte, T., Hutten, R.C.B., Visser, R.G.F., and van Eck, H.J.** (2011). Towards F1 Hybrid Seed Potato Breeding. *Potato Res.* **54**: 301–312.

- Liu, C. et al.** (2017). A 4-bp Insertion at ZmPLA1 Encoding a Putative Phospholipase A Generates Haploid Induction in Maize. *Mol. Plant* **10**: 520–522.
- Maheshwari, S., Tan, E.H., West, A., Franklin, F.C.H., Comai, L., and Chan, S.W.L.** (2015). Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genet.* **11**: e1004970.
- Marks, G.E.** (1966). The enigma of triploid potatoes. *Euphytica* **15**: 285–290.
- Miller, J.C.** (1954). Selection of desirable somatic mutations; A means of potato improvement. *Am. Potato J.* **31**: 358–359.
- Miller, J.C., Scheuring, D.C., Miller, J.P., and Fernandez, G.C.J.** (1999). Selection, evaluation, and identification of improved Russet Norkotah strains. *Am. J. Potato Res.* **76**: 161–167.
- Miyao, A., Nakagome, M., Ohnuma, T., Yamagata, H., Kanamori, H., Katayose, Y., Takahashi, A., Matsumoto, T., and Hirochika, H.** (2012). Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. *Plant Cell Physiol.* **53**: 256–264.
- Neelakandan, A.K. and Wang, K.** (2012). Recent progress in the understanding of tissue culture-induced genome level changes in plants and potential applications. *Plant Cell Rep.* **31**: 597–620.
- Ordoñez, B., Santayana, M., Aponte, M., Henry, I.M., Comai, L., Eyzaguirre, R., Lindqvist-Kreuzer, H., and Bonierbale, M.** (2021). PL-4 (CIP596131.4): an Improved Potato Haploid Inducer. *Am. J. Potato Res.* **98**: 255–262.
- Park, J.-S., Park, J.-H., Kim, S.-J., and Park, Y.-D.** (2020). Genome analysis of tissue culture-derived variations in regenerated *Brassica rapa* ssp. *pekinensis* plants using next-generation sequencing. *Horticulture, Environment, and Biotechnology* **61**: 549–558.
- Peschke, V.M. and Phillips, R.L.** (1991). Activation of the maize transposable element Suppressor-mutator (Spm) in tissue culture. *Theor. Appl. Genet.* **81**: 90–97.
- Peschke, V.M., Phillips, R.L., and Gengenbach, B.G.** (1987). Discovery of Transposable Element Activity Among Progeny of Tissue Culture—Derived Maize Plants. *Science*.
- Pham, G.M., Braz, G.T., Conway, M., Crisovan, E., Hamilton, J.P., Laimbeer, F.P.E., Manrique-Carpintero, N., Newton, L., Douches, D.S., Jiang, J., Veilleux, R.E., and Buell, C.R.** (2019). Genome-wide Inference of Somatic Translocation Events During Potato Dihaploid Production. *Plant Genome* **12**.
- Pham, G.M., Hamilton, J.P., Wood, J.C., Burke, J.T., Zhao, H., Vaillancourt, B., Ou, S., Jiang, J., and Buell, C.R.** (2020). Construction of a chromosome-scale long-read reference genome assembly for potato. *Gigascience* **9**.
- Pham, G.M., Newton, L., Wiegert-Rininger, K., Vaillancourt, B., Douches, D.S., and Buell, C.R.** (2017). Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. *Plant J.*
- Phillips, R.L., Kaepler, S.M., and Olhoft, P.** (1994). Genetic instability of plant tissue cultures: breakdown of normal controls. *Proc. Natl. Acad. Sci. U. S. A.* **91**: 5222–5226.
- Prodhomme, C., Vos, P.G., Paulo, M.J., Tammes, J.E., Visser, R.G.F., Vossen, J.H., and van Eck,**

- H.J.** (2020). Distribution of P1(D1) wart disease resistance in potato germplasm and GWAS identification of haplotype-specific SNP markers. *Theor. Appl. Genet.* **133**: 1859–1871.
- Pucker, B., Rückert, C., Stracke, R., Viehöver, P., Kalinowski, J., and Weisshaar, B.** (2019). Twenty-Five Years of Propagation in Suspension Cell Culture Results in Substantial Alterations of the *Arabidopsis Thaliana* Genome. *Genes* **10**.
- Ramulu, K.S., Sree Ramulu, K., Dijkhuis, P., and Roest, S.** (1983). Phenotypic variation and ploidy level of plants regenerated from protoplasts of tetraploid potato (*Solanum tuberosum* L. cv. “Bintje”). *Theoretical and Applied Genetics* **65**: 329–338.
- Ravi, M. and Chan, S.W.L.** (2010). Haploid plants produced by centromere-mediated genome elimination. *Nature* **464**: 615–618.
- Sanei, M., Pickering, R., Kumke, K., Nasuda, S., and Houben, A.** (2011). Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. *Proc. Natl. Acad. Sci. U. S. A.* **108**: E498–505.
- Sato, M., Kawabe, T., Hosokawa, M., Tatsuzawa, F., and Doi, M.** (2011). Tissue culture-induced flower-color changes in *Saintpaulia* caused by excision of the transposon inserted in the flavonoid 3', 5' hydroxylase (F3'5'H) promoter. *Plant Cell Rep.* **30**: 929–939.
- Seymour, D.K., Filiault, D.L., Henry, I.M., Monson-Miller, J., Ravi, M., Pang, A., Comai, L., Chan, S.W.L., and Maloof, J.N.** (2012). Rapid creation of *Arabidopsis* doubled haploid lines for quantitative trait locus mapping. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 4227–4232.
- Sharma, S.K., MacKenzie, K., McLean, K., Dale, F., Daniels, S., and Bryan, G.J.** (2018). Linkage Disequilibrium and Evaluation of Genome-Wide Association Mapping Models in Tetraploid Potato. *G3* **8**: 3185–3202.
- Stelpflug, S.C., Eichten, S.R., Hermanson, P.J., Springer, N.M., and Kaeppler, S.M.** (2014). Consistent and heritable alterations of DNA methylation are induced by tissue culture in maize. *Genetics* **198**: 209–218.
- Straadt, I.K. and Rasmussen, O.S.** (2003). AFLP analysis of *Solanum phureja* DNA introgressed into potato dihaploids. *Plant Breed.* **122**: 352–356.
- Stroud, H., Ding, B., Simon, S.A., Feng, S., Bellizzi, M., Pellegrini, M., Wang, G.-L., Meyers, B.C., and Jacobsen, S.E.** (2013). Plants regenerated from tissue culture contain stable epigenome changes in rice. *Elife* **2**: e00354.
- Sun, H., Jiao, W.B., Krause, K., Campoy, J.A., and Goel, M.** (2021). Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *bioRxiv*.
- Tan, E.H., Henry, I.M., Ravi, M., Bradnam, K.R., Mandakova, T., Marimuthu, M.P., Korf, I., Lysak, M.A., Comai, L., and Chan, S.W.** (2015). Catastrophic chromosomal restructuring during genome elimination in plants. *Elife* **4**.
- Tang, X. et al.** (2018). A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biol.* **19**: 84.
- Veilleux, R.E. and Johnson, A.A.T.** (2010). Somaclonal variation: Molecular analysis, transformation interaction, and utilization. In *Plant Breeding Reviews* (John Wiley & Sons, Inc.: Oxford, UK), pp.

229–268.

- Vining, K., Pomraning, K.R., Wilhelm, L.J., Ma, C., Pellegrini, M., Di, Y., Mockler, T.C., Freitag, M., and Strauss, S.H.** (2013). Methylome reorganization during in vitro dedifferentiation and regeneration of *Populus trichocarpa*. *BMC Plant Biol.* **13**: 92.
- Wangenheim, K.H., Peloquin, S.J., and Hougas, R.W.** (1960). Embryological investigations on the formation of haploids in the potato (*Solanum tuberosum*). *Molecular and General.*
- Wang, N., Xia, X., Jiang, T., Li, L., Zhang, P., Niu, L., Cheng, H., Wang, K., and Lin, H.** (2021). In planta haploid induction by genome editing of DMP in the model legume *Medicago truncatula*. *Plant Biotechnol. J.*
- Waugh, R., Baird, E., and Powell, W.** (1992). The use of RAPD markers for the detection of gene introgression in potato. *Plant Cell Rep.* **11**: 466–469.
- Wheeler, V.A., Evans, N.E., Foulger, D., Webb, K.J., Karp, A., Franklin, J., and Bright, S.W.J.** (1985). Shoot Formation from Explant Cultures of Fourteen Potato Cultivars and Studies of the Cytology and Morphology of Regenerated Plants. *Ann. Bot.* **55**: 309–320.
- Wilkinson, M.J., Bennett, S.T., Clulow, S.A., Allainguillaume, J., Harding, K., and Bennett, M.D.** (1995). Evidence for somatic translocation during potato dihaploid induction. *Heredity* **74 (Pt 2)**: 146–151.
- Yao, L., Zhang, Y., Liu, C., Liu, Y., Wang, Y., Liang, D., Liu, J., Sahoo, G., and Kelliher, T.** (2018). OSMATL mutation induces haploid seed formation in indica rice. *Nat Plants* **4**: 530–533.
- Ye, M., Peng, Z., Tang, D., Yang, Z., Li, D., Xu, Y., Zhang, C., and Huang, S.** (2018). Generation of self-compatible diploid potato by knockout of S-RNase. *Nat Plants* **4**: 651–654.
- Zhang, C., Wang, P., Tang, D., Yang, Z., Lu, F., Qi, J., Tawari, N.R., Shang, Y., Li, C., and Huang, S.** (2019). The genetic basis of inbreeding depression in potato. *Nat. Genet.* **51**: 374–378.
- Zhang, C., Yang, Z., Tang, D., Zhu, Y., Wang, P., Li, D., Zhu, G., Xiong, X., Shang, Y., Li, C., and Huang, S.** (2021). Genome design of hybrid potato. *Cell.*
- Zhang, D. et al.** (2014). Tissue culture-induced heritable genomic variation in rice, and their phenotypic implications. *PLoS One* **9**: e96879.
- Zhong, Y. et al.** (2020). A DMP-triggered in vivo maternal haploid induction system in the dicotyledonous *Arabidopsis*. *Nat Plants* **6**: 466–472.
- Zhong, Y. et al.** (2021). A genotype independent DMP-HI system in dicot crops. *bioRxiv*: 2021.06.21.449224.
- Zhong, Y. et al.** (2019). Mutation of ZmDMP enhances haploid induction in maize. *Nat Plants* **5**: 575–580.
- Zhou, Q., Tang, D., Huang, W., Yang, Z., Zhang, Y., Hamilton, J.P., Visser, R.G.F., Bachem, C.W.B., Robin Buell, C., Zhang, Z., Zhang, C., and Huang, S.** (2020). Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nat. Genet.* **52**: 1018–1023.

Chapter 2

Genomic Outcomes of Haploid Induction Crosses in Potato (*Solanum tuberosum* L.)

[Published in: Genetics 214 (2): 369-380]

Kirk R Amundson¹, Benny Ordoñez^{1,2}, Monica Santayana², Ek Han Tan^{1,3}, Isabelle M Henry¹, Elisa Mihovilovich², Merideth Bonierbale^{2,4}, and Luca Comai^{1*}

¹Genome Center, University of California, Davis, 1 Shields Avenue, Davis, CA, 95616, USA.

²International Potato Center (CIP), P.O. Box 1558, Lima 12, Peru.

³School of Biology and Ecology, University of Maine, Orono, ME 04469, USA.

⁴Present Address: Duquesa Business Centre, P. O. Box 157, Manilva, Malaga, Spain 29692.

*Corresponding author: lcomai@ucdavis.edu

Short title: Parental contribution to haploids

The author responsible for distribution of materials integral to the findings presented in the article in accordance with the policy described in the Instructions for Authors is: Luca Comai (lcomai@ucdavis.edu)

Abstract

The challenges of breeding autotetraploid potato (*Solanum tuberosum*) have motivated the development of alternative breeding strategies. A common approach is to obtain uniparental dihaploids from a tetraploid of interest through pollination with *S. tuberosum* Andigenum Group (formerly *S. phureja*) cultivars. The mechanism underlying haploid formation of these crosses is unclear, and questions regarding the frequency of paternal DNA transmission remain. Previous reports described aneuploid and euploid progeny, which, in some cases, displayed genetic markers from the haploid inducer. Here, we surveyed a population of 167 presumed dihaploids for large-scale structural variation that would underlie chromosomal addition from the haploid inducer, and for small-scale introgression of genetic markers. In 19 progeny, we detected ten of the twelve possible trisomies and, in all cases, demonstrated the non-inducer parent origin of the additional chromosome. Deep sequencing indicated that occasional, short-tract signals appearing of haploid inducer origin were better explained as technical artifacts. Leveraging recurring CNV patterns, we documented sub-chromosomal dosage variation indicating segregation of polymorphic maternal haplotypes. Collectively, 52% of assayed chromosomal loci were classified as dosage variable. Our findings help elucidate the genomic consequences of potato haploid induction and suggest that most potato dihaploids will be free of residual pollinator DNA.

Summary

The cultivated potato's polyploid and highly heterozygous genome is challenging to manipulate by breeding. Genome reduction to diploidy can be achieved by pollinating tetraploids with specialized varieties called haploid inducers. These crosses result in individuals with two

chromosome sets from the seed parent. It is unclear how much, if any, of the pollinator genome is retained by these diploids. Here we searched the genomes of 167 such diploids for residual pollinator DNA. While copy number changes and aneuploidy were frequent, they were all attributable to maternal DNA. We conclude that pollinator contributions are, at most, very rare in this population.

Introduction

Highly prized in plant breeding and research, haploid plants can be obtained through culture of immature gametophytes or, more conveniently, through inter- or intraspecific crosses in which the genome of one parent, the haploid inducer (HI), does not appear in the progeny (Ishii et al., 2016a; Forster et al., 2007a). Having been documented in 74 crosses between monocotyledonous species and 35 involving dicotyledonous species, this phenomenon is not uncommon (reviewed in (Ishii et al., 2016a)), and is widely exploited for the rapid generation of inbred lines, as well as genetic mapping and germplasm base expansion. The genetic properties that make a haploid inducer, however, are largely unknown with a couple of exceptions. Artificial manipulation of centromeric histone H3 can result in a haploid inducer (Ravi and Chan, 2010; Maheshwari et al., 2015; Kuppu et al., 2015; Kelliher et al., 2016; Karimi-Ashtiyani et al., 2015). Furthermore, natural maize haploid inducers depend on inactivation of the phospholipase encoded by the Matrilineal locus (Kelliher et al., 2017; Gilles et al., 2017; Liu et al., 2017).

In potato (*Solanum tuberosum* L.), the world's fourth most important crop in terms of calories consumed per person per day (<http://www.fao.org/faostat/en/#compare>), haploid seed can be routinely obtained via pollination with select haploid inducer varieties from the diploid *S.*

tuberosum Andigenum Group (formerly *S. tuberosum* Phureja Group or *S. phureja* (Spooner et al., 2014) . Such crosses with tetraploid potato ($2n=4x=48$) produce $2n=2x=24$ dihaploids that can be used for genetic mapping (Mihovilovich et al., 2014; Velásquez et al., 2007; Ercolano et al., 2004; Kotch et al., 1992; Pineda et al., 1993; Bartkiewicz et al., 2018). Additionally, these crosses produce hybrids that can be either triploid or tetraploid (Hanneman and Ruhde, 1978; Wagenvoort and Lange, 1975), and can be identified as seed because they express a purple embryo spot, a dominant anthocyanin marker encoded by the haploid inducers that is expected to be absent in the dihaploids (Fig. 2.1) (Hermsen and Verdenius, 1973a). In embryo spot-negative dihaploid populations, 3.5-11.0% aneuploids are commonly found, exhibiting $2n=2x+1=25$, and rarely, $2n=2x+2=26$ karyotypes (Wagenvoort and Lange, 1975).

An ongoing question regarding haploid induction in potato is the cytogenetic mechanism by which it occurs. Two mechanisms have been proposed. The first mechanism is parthenogenesis, in which haploid inducer pollen triggers the development of unfertilized egg cells without making a genetic contribution to the embryo. This is supported by three lines of evidence: i) endosperms from $4x$ by $2x$ potato haploid induction crosses are usually hexaploid instead of the expected pentaploid, suggesting abnormal pollen (Wangenheim et al., 1960); ii) haploid inducers frequently produce 24-chromosome restitution sperm nuclei, thought to be a consequence of failed pollen mitosis II; iii) colchicine-treated pollen of non-inducer *S. tarjinse* also exhibit restitution sperm nuclei and can induce haploids (Montelongo-Escobedo and Rowe, 1969). Based on these observations, it was speculated that a $2x$ restitution sperm fertilizes the central cell, leaving no sperm to fertilize the egg. The second mechanism is genome elimination, in which haploid inducer chromosomes are eliminated from the embryo after fertilization. This

alternative hypothesis is supported by the presence of inducer-specific AFLP, RFLP, or isozyme markers in presumably dihaploid progeny. Often, progeny exhibiting genetic markers from the haploid inducer are also aneuploid, suggesting inheritance of an entire chromosome from the haploid inducer (Clulow et al., 1991; Waugh et al., 1992; Clulow et al., 1993; Wilkinson et al., 1995; Clulow and Rousselle-Bourgeois, 1997; Straadt and Rasmussen, 2003; Ercolano et al., 2004; Allainguillaume et al., 1997). These results are consistent with haploid induction crosses in maize (Riera-Lizarazu et al., 1996; Zhao et al., 2013), Arabidopsis (Tan et al., 2015; Kuppu et al., 2015; Maheshwari et al., 2015), and oat-maize hybrids (Riera-Lizarazu et al., 1996) in which one or more haploid inducer chromosomes persist in otherwise haploid plants.

Recently, widespread and ubiquitous introgression of very short DNA regions (>100bp) from the haploid inducer genome into potato dihaploids has been reported by SNP genotyping (Bartkiewicz et al., 2018) and by whole genome sequencing (Pham et al., 2019). In the latter case, depending on the progeny, 25,000 to 300,000 translocation events were inferred, suggesting a massive contribution from the transient haploid inducer genome to the maternally contributed genome. Genetic information in short segments of HI DNA could persist through three mechanisms: i) non homologous recombination leading, for example, to insertion; ii) homologous recombination leading, for example, to gene conversion, and iii) autonomous replication. To clarify the underlying molecular arrangement, we use the term “addition” to indicate the presence in the dihaploid genome of an additional copy derived from the HI. We use the term “introgression” to indicate the DNA from the HI has recombined with the donor genome. If recombination is homologous, this could result in copy-neutral transfer of information. If confirmed, such widespread recombination would require rethinking of both

breeding and biotechnology experimental strategies to either avoid or exploit it, depending on context.

In light of these observations, we resequenced a population of 167 primary dihaploids derived from tetraploid Andigenum Group cultivar Alca Tarma to address three questions: First, does a curated set of phenotypically normal primary dihaploids display aneuploidy? If so, which parent contributes the additional chromosome(s)? Second, are single chromosomes or large chromosome segments from the haploid inducer added to otherwise dihaploid progeny? Third, are shorter segments of the haploid inducer genome introgressed or added to the dihaploids? If so, on what scale? We considered two hypotheses: first, that occasional failure to eliminate the entire HI chromosome set could result in persistence of an additional chromosome, whole or fragmentary, in an otherwise dihaploid potato. We have previously demonstrated that events involving entire chromosomes or chromosome segments are readily detectable with low coverage whole genome sequencing and chromosome dosage analyses in *Arabidopsis* (Henry et al., 2010; Maheshwari et al., 2015; Tan et al., 2015; Kuppu et al., 2015) and poplar (Henry et al., 2015; Zinkgraf et al., 2017). In potato, extensive copy number variation (CNV), which has been described in numerous cytological and genomic surveys of diploid and tetraploid cytotypes, is a potentially confounding factor that should be taken into account (Iovene et al., 2013; de Boer et al., 2015; Hardigan et al., 2016; Pham et al., 2017; Hardigan et al., 2017). Second, smaller scale introgression described could be detected by deeper sequencing of selected individuals. Our genomic analysis did reveal frequent whole-chromosome aneuploidy and widespread segmental dosage variation, but these were never attributable to the haploid inducer. Notwithstanding the ability of dihaploids to tolerate chromosomal dosage imbalance, we found no evidence that

haploid inducers contributed large chromosomal segments to the progeny. Further, by a set of standard criteria, we found no short segmental introgression either.

Results

Induction, selection and sequencing of dihaploids

A population of 167 primary dihaploids was generated from Alca Tarma via pollination with haploid inducers IVP101 or PL4 (Velásquez et al., 2007; Mihovilovich et al., 2014) over the course of two previous studies (Velásquez et al., 2007; Mihovilovich et al., 2014). The dihaploids in this population lacked the homozygous dominant embryo spot seed marker present in both haploid inducers and displayed the expected count of guard cell chloroplast counts and root cell chromosomes (Velásquez et al., 2007; Mihovilovich et al., 2014) (Fig. 2.1). We carried out genome resequencing to identify dosage variation and aneuploidy among the population. Alca Tarma, IVP101, PL4 and three selected haploids were sequenced to 40-56x coverage. For the remaining dihaploids, we generated an average of 3.88 million reads per individual (Supplemental Table S2.1).

Maternally derived trisomy

We hypothesized that introgressions into the host genome could derive from at least three types of events, each associated with specific predictions: i) non-homologous transposition of haploid inducer (HI) segments to the host genome, resulting in three copies of the corresponding region with a SNP ratio of 1 HI : 2 host; ii) homologous recombination leading to replacement of a segment, either interstitial or terminal, resulting in no copy number change of the affected region and a SNP ratio of 1 HI : 1 host; iii) gene conversion from a non-crossover event producing a

very short (25-50bp) conversion tract and resulting mostly in a single SNP with a 1:1 ratio. Formally, homologous recombination could also result in duplication and resemble case i) (Das et al., 1991).

We first screened the population for whole-chromosome aneuploidy. Sequencing reads were aligned to the DM1-3 reference genome, and read counts per chromosome were normalized to those of the tetraploid parent such that values near 2.0 were obtained for chromosomes present in two copies, and values deviating from 2.0 indicated aneuploidy. In 19 individuals, standardized coverage was significantly elevated for a single chromosome, suggesting a primary trisomy (Fig. 2.2A). Root tip chromosome spreads were evaluated for 15 of the 19 putative trisomics, confirming the $2n=2x+1=25$ karyotype in all cases (Fig. 2.2B-C; Supplemental Fig. S2.1). In total, six trisomics of chromosome 2, two of chromosomes 4, 5, 7, and 8, and one of chromosomes 1, 3, 6, 10, and 12 were detected in the population (Fig. 2.2A; Supplemental Fig. S2.1). To determine the parental origin of each trisomy, 382,967 SNPs homozygous for the same allele in both HIs, but homozygous for an alternate allele in Alca Tarma were identified from sequencing the parental genomes (Supplemental Data Set S2.1). The fraction of haploid inducer-specific allele calls along all chromosomes was then calculated for each trisomic individual. Using this measurement, a trisomy from either haploid inducer was expected to exhibit approximately 33% haploid inducer allele across the affected chromosome. To empirically validate this expectation, we evaluated 200 simulated low-coverage hybrids each consisting of 2 million reads from Alca Tarma and 1 million reads from either IVP101 or PL4 (see Methods). In each trisomic dihaploid, HI alleles were nearly absent from the trisomic chromosome (Fig. 2.2D), indicating inheritance from the non-inducer parent.

Determining which SNP bins are informative

To survey the population for haploid inducer chromosome addition or introgression, the SNP dosage analysis described above was repeated using non-overlapping 1Mb bins. To account for low density of homozygous parental SNP markers in some regions, we included an additional set of SNPs that did not fit the optimal criteria of the original set (Supplemental Data Set S2.2). Of the added SNPs, 170,273 were heterozygous in one haploid inducer and homozygous in the other while 247,144 were heterozygous in both haploid inducers. A consequence of including these additional SNP markers is that a haploid inducer allele contribution could be lower than the expected 33%. Therefore, we empirically determined the expected percentages for each bin by comparing the percentages obtained from low-coverage *in silico* triploid hybrids and negative controls. Any bin in which the distributions of observed HI allele percentages of the hybrid and negative control groups exhibited any overlap, was withheld from consideration. To call an introgression, we required at least three adjacent bins to exhibit a haploid inducer allele percentage that overlapped with the empirical thresholds determined from the *in silico* hybrid analysis. Among 1Mb bins that were considered in this analysis, no such events were found (Supplemental Fig. S2.2). Notably, a recombination event that substitutes an Alca Tarma chromosome segment by the corresponding segment from either haploid inducer would produce a higher HI allele percentage than an addition event (50% vs 33%), suggesting that neither addition nor introgression of a haploid inducer chromosome segment occurred in the dihaploid population.

CNV analysis

To complement the approach described above, we also asked whether rare structural variants existed in the population and if so, from which parent they were derived. The dosage analysis was repeated as described above for non-overlapping 250kb bins of the reference genome. As before, a value of 2 indicates the expected diploid complement and values deviating from 2 indicate structural variation. Inferred karyotypes of a representative $2n=2x=24$ dihaploid and a $2n=2x+1=25$ maternal trisomy are shown in Figure 2.2C and 2.2D, respectively. By overlaying dosage plots for each dihaploid, it became evident that many CNVs are recurrent in the population, with copy number gains and losses of the same locus among the dihaploids (Fig. 2.3A, .23B; Supplemental Fig. S2.3). This behavior is evident in the pericentromeric heterochromatic region of all chromosomes, while the euchromatic, gene-rich arms are more uniform (Fig. 2.3, A-G). Chromosomes 2 and 4 are exceptions in that they display recurring dosage variation in their short arms (Fig. 2.3, B; Supplemental Fig. S2.3).

To define structurally polymorphic loci and alleles at these loci, we clustered relative coverage values separately for each 250kb bin (Supplemental Fig. S2.4). Structural variation was widespread, with multiple clusters detected in 48% of 250kb bins (Fig. 2.3G). From joint consideration of the number of individuals in a cluster and read depth of high-coverage samples, we inferred that most dosage variants represented segregating polymorphism among Alca Tarma haplotypes, as exemplified by a 1 Mb block of chromosome 6 that is present on one chromosome and absent on the other three (Fig. 2.4, Supplemental Fig. S2.5). Among 14 duplications that were ≥ 750 kb and present in $<5\%$ of dihaploids, five were clearly derived from Alca Tarma (Supplemental Table 2). Based on comparison of simulated hybrids, SNP marker density was too

low to conclusively resolve the parental origin of the remaining nine, but the fraction of HI SNP in power analyses classified them as low probability outliers (Supplemental Fig. S2.6).

SNP loci consistent with addition or introgression were rare and dispersed

To search for addition and introgression events at higher resolution, three dihaploids were sequenced to 19-30x depth with Illumina paired-end sequencing. To minimize spurious genotype calls, SNP loci that passed our quality filtering steps were further filtered to exclude sites with even a single Alca Tarma read that matched a haploid inducer allele. In total, 800,384 loci were assayed, with 725,952-745,535 loci assayed in each of the three dihaploids. The fraction of loci with heterozygous genotypes consistent with addition or introgression was very low (0.157-0.195%). Among heterozygous sites in the dihaploids, haploid inducer alleles tended to be underrepresented relative to expectations for either addition (~33%) or introgression (50%) of haploid inducer DNA (Supplemental Fig. S2.7), and read depth was lower in both Alca Tarma and the dihaploid at hand (Supplemental Fig. S2.8). Upon observing the low read depth and underrepresentation of haploid inducer alleles at many putative introgression loci, we applied additional filters (minimum haploid inducer allele depth ≥ 5 ; minimum allele depth representation 10%) to investigate the subset of possible introgression events with the best read support in our dataset. Applying these filters reduced the number of putative introgression loci from 1,217 to 266 in LOP868.004, from 1,457 to 358 in LOP868.064, and from 1,124 to 198 in LOP868.305 (Supplemental Data Set S2.3).

We further investigated the distribution of all putative introgression loci with respect to the reference genome. As loci that were heterozygous in the two haploid inducers were also included in this analysis and the phase is unknown, introgression of a haploid inducer segment may appear

discontinuous. Therefore, we estimated a lower bound of the number of introgression events using a seed-and-extend approach: starting at SNP loci with inducer-specific alleles introgression events, putative introgressions were extended in both directions until a parent-homozygous SNP locus with no evidence of haploid inducer alleles in the dihaploid was encountered. The total number of SNP markers, as well as markers exhibiting haploid inducer alleles were tallied for each event. Dihadpoids LOP868.004, LOP868.064 and LOP868.305 exhibited 1,037, 1,191 and 806 events, respectively. Approximately half (49.6%) of putative introgression events consisted of a singleton SNP, i.e., a single locus exhibiting haploid inducer alleles flanked by homozygous parental loci that did not support introgression in the dihaploid (mean distance between loci flanking a singleton = 9,110 bp; median distance = 735bp). Among non-singleton events, very few markers exhibited haploid inducer alleles for each event (Supplemental Fig. S2.9).

Finally, read alignments were manually inspected for the introgression event with the highest number of parental SNP loci exhibiting haploid inducer alleles. This event spans a ~2Mb region of chromosome 5 in dihaploid LOP868.305, including 5,712 parental SNP loci, of which only 43 exhibited haploid inducer alleles. Specifically, we looked for phased variants on the same read consistent with introgression of a contiguous haploid inducer haplotype. Remarkably, no reads supporting haploid inducer introgression at each of these 43 loci matched a local haplotype present in either haploid inducer (Supplemental Fig. S2.10, Supplemental Data Set S2.4).

Taken together, these analyses revealed rare maternally associated chromosome remodeling among a backdrop of widespread structural heterogeneity (Fig. 2.3A), including several large and novel variants detected in the genome of the tetraploid parent. Robust evidence for

chromosomal introgression from the haploid inducer was not detected in any case, and a detailed survey of three dihaploid genomes revealed sites that, while superficially consistent with introgression of haploid inducer DNA, resemble sequencing or alignment artifacts.

Discussion

We analyzed the genomes of 167 primary dihaploids produced by pollinating the *S. tuberosum* Andigenum Group cultivar Alca Tarma with haploid inducers IVP101 and PL4. This population is representative of a typical dihaploid progeny set used for breeding in that, during its development, selection has been applied against individuals with DNA content differing from the dihaploid state, against individuals carrying the genetic color marker from the haploid inducer (Fig. 2.1), and against severe abnormality. Using low-pass sequencing, we derived karyotypes of each progeny, identifying primary trisomy in 11.4% of individuals. By comparing parental SNP genotypes, we established that whole-chromosome aneuploidy was maternally inherited.

Widespread variation in DNA dosage consistent with segregation of maternal structural variation was already evident at 1Mb resolution (Fig. 2.3). This analysis does not consider small structural events on genic or transposable element scale. Using randomly downsampled data to simulate a triploid hybrid, we empirically show that our low-pass sequencing approach provided an effective and affordable method to decipher the complexity of populations with highly variable genomic structure, such as often employed in breeding (Barrell et al., 2013; Hirsch et al., 2013). Our findings lead to several conclusions.

Genetic haploid inducers act by either stimulating parthenogenesis in the female or chromosome instability in the embryo resulting in missegregation and loss of one parental chromosome set. In

potato, evidence in support of parthenogenesis has been reported (Wangenheim et al., 1960; Montelongo-Escobedo and Rowe, 1969; Peloquin et al., 1996). At the same time, genome elimination is supported by the detection of genetic markers from the haploid inducer in euploids and aneuploids arising from haploid induction crosses (Clulow et al., 1991; Waugh et al., 1992; Clulow et al., 1993; Wilkinson et al., 1995; Clulow and Rousselle-Bourgeois, 1997; Allainguillaume et al., 1997; Straadt and Rasmussen, 2003; Ercolano et al., 2004). Whole genome sequencing provides a more informative and reliable method to assess the genetic contribution of the haploid inducer. In *Arabidopsis* haploids, DNA from the haploid inducer can be identified readily from low-pass sequencing (Tan et al., 2015). It consists of whole chromosomes or segmental subsets of a single chromosome, consistent with the incomplete elimination of certain chromosomes, which persist autonomously, whole or rearranged.

We employed a similar approach with the dihaploids derived from Alca Tarma. Chromosomal addition or introgression comparable to described cases should be evident by the appearance of whole chromosomes or large segments containing a haploid inducer centromere (Riera-Lizarazu et al., 1996; Tan et al., 2015; Zhao et al., 2013). Although aneuploidy was common, convincing evidence of chromosomal contribution from the haploid inducers was not observed, consistent with AFLP analysis on this set (Velásquez et al., 2007) and analysis of another dihaploid population (Samitsu and Hosaka, 2002). The capability of our method to identify long chromosomal segments that diverge in SNP or copy number, is validated by *in silico* reconstructions (Methods) and effectiveness in comparable systems (Tan et al., 2015; Henry et al., 2015), indicating that the transfer of large segments of haploid inducer DNA (Zhao et al., 2013; Riera-Lizarazu et al., 1996; Tan et al., 2015; Kuppu et al., 2015) did not take place.

Assessment of small introgressions is more challenging. Recent genotyping (Bartkiewicz et al., 2018) or sequencing (Pham et al., 2019) of other dihaploid potato populations found that ~1% of SNP loci displayed presence of HI DNA in very small tracts and with lower than expected allelic ratio. As in these reports, using high coverage sequence data from three dihaploids we detected many, widely dispersed SNP represented by proportionally fewer aligned reads than expected for addition of a haploid inducer DNA segment, or alternatively, replacement of a non-HI haplotype by homologous recombination. To address these observations, Pham et al. suggested that tens of thousands of recombination events took place in each dihaploid during early growth before the haploid inducer genome was eliminated. Further, they proposed that low allelic frequency could be explained by tissue chimerism. Mechanistically, this type of short introgression could be explained by somatic recombination caused by double stranded DNA breaks followed by synthesis-dependent strand annealing (SDSA) or dsDNA break repair (DSBR) (Pâques and Haber, 1999). Notably, while recombination of ectopic sequences has been demonstrated in plants (Čermák et al., 2017; Filler Hayut et al., 2017; Puchta, 1999), these events are infrequent and require careful interpretation (Puchta and Hohn, 2012). In this case, the scale of these changes ranged from ~30,000 to 300,000 per sequenced haploid and affected all examined haploids (Pham et al., 2019), implying extremely high efficiency of recombination. The hypothesis of autonomous replication of HI DNA segments, for which there are precedents (Cohen et al., 2008; Shibata et al., 2012), could relieve the need for recombination. Nevertheless, it would also require high efficiency propagation of extrachromosomal elements. These problems suggest a conservative interpretation of our data: these signals are artifactual and could have

causes comparable to those identified in other gene conversion studies (Wijnker et al., 2013; Qi et al., 2014).

We conclude that, for the cross between haploid inducers IVP101 or PL4 and tetraploid Alca Tarma, either the mechanism of haploid induction did not involve egg fertilization, or genome elimination resulted in loss of all haploid inducer chromosomes before the plants were evaluated. Events resulting in chromosome addition or introgression may be infrequent. For this reason, it may be premature to rule out genome elimination until more dihaploids derived from different parental combinations are evaluated. If haploid induction acts via genome elimination, both the addition of large DNA segments in the form of chromosomes and the rare introgression of small segments via recombination could be identified and used for manipulation of the potato genome.

Our findings suggest that after tuberosum x phureja crosses, plants derived from seeds that did not express the purple spot marker and that display $2x$ genome content by flow cytometry are likely to be clean dihaploids (free of pollinator genome). That 11.4% of the dihaploid population evidently escaped initial screening against aneuploidy based on chloroplast counts, visual phenotyping, and chromosome counts underscores the difficulty of identifying aneuploids in highly variable dihaploid progeny. Their occurrence is consistent with the high frequency of aneuploid gametes in autotetraploids (Comai, 2005) and with the ability of certain genotypes to tolerate imbalance (Rick and Notani, 1961; Henry et al., 2010). The frequency of chromosome 2 trisomy, which carries the nucleolar organizing region (NOR) of potato, may be explained if increasing the ribosomal RNA gene copy number offsets the disadvantage of linked dosage

imbalance. Alternatively, ribosomal gene transcription (or other unknown feature) could interfere with segregation (Tomson et al., 2006).

Conclusions

We undertook this investigation with the objective to assess large-scale structural variation and its causes in dihaploids produced through genetic induction. Using cost effective low pass sequencing, we documented extensive, large-scale structural variation affecting over 52% of the genome. We found that 11% of the dihaploids were trisomic, frequently for the chromosome that carries the nucleolar-organizing region. In spite of multiple previous reports of genomic contamination by the haploid inducer used as pollinator, we did not detect any large-scale introgression of the haploid inducer genome.

Materials and Methods

Plant material

A population of 167 putative dihaploids described in Velasquez et al. (2007) and Mihovilovich et al. (2014) was raised from the progeny of tetraploid Andigenum Group landrace cultivar Alca Tarma and one of two haploid inducer genotypes: IVP-101 or CIP596131.4. For convenience, we refer to CIP596131.4 as PL4 throughout. Both haploid inducers are homozygous for a dominant embryo spot marker that results in anthocyanin accumulation at the base of the cotyledons visible through the seed coat. Only seeds lacking the embryo spot marker were planted, and seedlings exhibiting more than an average of eight chloroplasts per guard cell

(minimum ten measured cells) were discarded. Seedlings were germinated on soil and maintained as *in vitro* cuttings thereafter.

Genomic DNA library preparation, sequencing, and pre-processing

Approximately 750ng of genomic DNA extracted from leaf tissue as previously described (Ghislain M., Zhang D. P., Herrera, M. R., 1999) was sheared to an average size of 300bp using a Covaris E-220 sonicator in a 50µl reaction volume using the following settings: 175W peak power, 10% duty factor, 200 cycles per burst, 50s treatment time, 4°C minimum temperature, 9°C maximum temperature. Genomic libraries were constructed using 375ng of sheared DNA and a KAPA Hyper Prep Kit (cat. no KK8504) with half-scale reactions, custom 8bp dual-indexed adapters, and library amplification cycles as specified in Supplemental Table S1.

Libraries were sequenced on an Illumina HiSeq 4000 in either 50nt single-end or 150nt paired-end mode by the University of California, Davis DNA Technologies Core and Vincent Coates Genome Sequencing Laboratory.

Libraries were demultiplexed using a custom Python script available on our lab website (allprep-12.py; http://comailab.genomecenter.ucdavis.edu/index.php/Barcoded_data_preparation_tools).

Variant calling

For paired-end sequencing, sequence reads were processed with Cutadapt (v.1.15) to remove low-quality (<Q10) bases, adapter sequences, and reads ≤40nt after trimming. The DM1-3 v4.04 genome assembly, as well as DM1-3 chloroplast and mitochondrion sequences were retrieved from (http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml), concatenated, and used as the reference sequence for paired-end read alignment with BWA MEM (v.0.7.12-r1039) (Li, 2013) with mismatch penalty 6 and all other parameters left at the program default. PCR

duplicates were removed using Picard (v.2.14) MarkDuplicates, and only reads with mates mapping to the same chromosome were retained. For reads with overlapping mates, one of the two reads was soft-clipped in the overlap region using bamUtil::clipOverlap (Jun et al., 2015). Variants were called on processed alignment files using freebayes (v.1.1.0) (Garrison and Marth, 2012) with minimum read mapping quality 41, minimum base quality 20, population priors not considered, and ploidy specified for each sample as a CNV map. To remove low-quality variants, the following site filters were applied in RStudio (v.3.4.0): NUMALT == 1, CIGAR == 1X, MQM \geq 50, MQMR \geq 50, |MQM - MQMR| < 10, RPPR \leq 20, RPP \leq 20, EPPR \leq 20, EPP \leq 20, SAP \leq 20, SRP \leq 20, DP \leq 344. Only sites with called homozygous Alca Tarma genotypes without reads matching haploid inducer alleles were retained. Sites that were called homozygous for the Alca Tarma allele in either haploid inducer were removed. Several additional quality filters were applied on each sample: depth \geq 10 in all three parents, \leq 5% Alca Tarma allele representation at called homozygous haploid inducer loci, and 40-60% Alca Tarma allele representation at called heterozygous haploid inducer loci. After filtering, 798,468 SNP were retained for analysis. Putative introgression loci were identified via heterozygous genotype calls in any of the three high-coverage dihaploids.

Dosage analysis

Single-end reads were aligned to the DM1-3 reference genome as described above, and only reads with mapping quality \geq Q10 were retained. Standardized coverage values were derived by taking the fraction of mapped reads that aligned to a given bin for that sample, dividing it by the corresponding fraction from the same bin for tetraploid LOP, and doubling the resulting value to indicate the expected diploid state. To mitigate mapping bias due to read type and length, Alca Tarma forward mates were hard-trimmed to 50 nt and remapped. When all chromosomes were

treated as equivalent, the distribution of per-chromosome standardized coverage values approximated a Gaussian distribution (QQ plots not shown), satisfying the assumption of a Z-score analysis. Aneuploidy was then called if chromosomal standardized coverage exceeded the all-chromosome population by ≥ 3 standard deviations. We identified local dosage variants by clustering standardized coverage values of non-overlapping 250kb bins using the R package MeanShift (CRAN - Package MeanShift). The clustering bandwidth parameter was set to the 50th percentile of inter-point distances in each 250kb bin.

Parental origin analyses of trisomy and dosage variants were carried out as previously described (Henry et al., 2015). Briefly, reads with mapping quality $\geq Q20$ and base calls $\geq Q20$ were used to compute allele-specific read depth at the subset of 800,384 SNP loci identified above located on chromosomes 1-12 of the DM1-3 v4.04 assembly. The percentage of reads supporting the haploid inducer allele reads among all reads at loci within a non-overlapping 1Mb bin was then reported. A biological positive control was not available for SNP analysis, we empirically evaluated limitations of the SNP dosage assay by comparing dihaploids with simulated triploid hybrids expected to resemble an introgressed haploid inducer chromosome segment at all tested genomic loci. To construct triploid hybrids *in silico*, pseudo-random subsets of exactly 2,015,413 and 1,007,706 forward mates were drawn 100 times from raw sequencing reads of Alca Tarma (SRA ID SRR6123032) and IVP101 (SRA ID SRR6123183), respectively. The number of parental reads was chosen such that parental reads would be present in a 2:1 ratio expected for a triploid, and that the number of raw reads for each *in silico* hybrid would match the 5th percentile of raw read counts in the dihaploid sequencing dataset. Similarly, 100 *in silico* hybrids were constructed using Alca Tarma and PL4 reads. As a negative control, pseudo-random subsets of

3023119 reads were drawn 100 times from Alca Tarma. Raw reads from all simulated hybrids were hard-trimmed to 50nt, and then processed using the SNP dosage pipeline described above with non-overlapping 1Mb bins. For each bin, if the ranges of haploid inducer allele percentage from either *in silico* hybrid group overlapped with the corresponding range of the negative control, the bin was withheld from analysis. Using this approach, 608 of 730 bins (83%) were considered in this analysis. Similarly, to determine whether unique dosage variants could be genotyped confidently, we compared the distributions of %HI allele values between simulated hybrid and negative control groups in the affected interval.

To estimate absolute copy number, per-position read was calculated using samtools depth, using only reads with mapping quality \geq Q20. Median read depth in non-overlapping 10kb windows was then determined using custom Python software available from ([github link](#)). We observed a positive correlation of window median read depth and GC content, suggesting PCR amplification bias introduced during sequencing library construction (Benjamini and Speed, 2012). For each 10kb bin, this bias was corrected by dividing median read depth by GC content for that bin. The resulting values were clustered using the MeanShift package in R. The centroid of the largest cluster was designated as the copy number corresponding to the expected ploidy (4 for Alca Tarma; 2 for dihaploids), and multiples of the centroid were used to designate the remaining copy number states.

Cytological analysis

Chromosome spreads were prepared from root tips as previously described (Watanabe and Orrillo, 1993), with minor modifications. Commercial permethrin was used at a concentration of 75 ppm as a pre-treatment to induce chromosome condensation. Roots were kept in an ice-cold

water bath for 24h before hydrolysis in 1N HCl for 10-15 minutes, then stained with lacto-propionic acid and squashed.

Data availability

Sequence data has been deposited in NCBI SRA BioProject ID PRJNA408137. Analysis code has been deposited at https://github.com/kramundson/LOP_manuscript. Supplemental material available at figshare: <https://doi.org/10.6084/m9.figshare.c.4714658>.

Funding

This work was supported by the National Science Foundation(NSF) (Plant Genome IOS Grant 1444612: Rapid and Targeted Intro-gression of Traits via Genome Elimination; to L.C.);

Acknowledgements

The authors would like to thank Awais Khan for encouraging and coordinating the early phase of this collaboration, Michell Feldmann, Jordan Weibel, Jeanine Montano and Helen Tsai for constructive comments on writing and data analysis, and the CGIAR Collaborative Program on Roots Tubers and Bananas (CRP – RTB) for supporting the development of the LOP population used here.

Author Contributions

KRA, EM, EHT, MB, IMH and LC designed experiments. MS performed cytological analysis. KRA, BO and EHT performed experiments. KRA and LC analyzed data and wrote the manuscript with input from all authors.

Literature Cited

- Allainguillaume, J., Wilkinson, M.J., Clulow, S.A., and Barr, S.N.R.** (1997). Evidence that genes from the male parent may influence the morphology of potato dihaploids. *Theor. Appl. Genet.* **94**: 241–248.
- Barrell, P.J., Meiyalaghan, S., Jacobs, J.M.E., and Conner, A.J.** (2013). Applications of biotechnology and genomics in potato improvement. *Plant Biotechnol. J.* **11**: 907–920.
- Bartkiewicz, A.M., Chilla, F., Terefe-Ayana, D., Lübeck, J., Strahwald, J., Tacke, E., Hofferbert, H.-R., Linde, M., and Debener, T.** (2018). Maximization of Markers Linked in Coupling for Tetraploid Potatoes via Monoparental Haploids. *Front. Plant Sci.* **9**: 620.
- Benjamini, Y. and Speed, T.P.** (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **40**: e72.
- de Boer, J.M., Datema, E., Tang, X., Borm, T.J.A., Bakker, E.H., van Eck, H.J., van Ham, R.C.H.J., de Jong, H., Visser, R.G.F., and Bachem, C.W.B.** (2015). Homologues of potato chromosome 5 show variable collinearity in the euchromatin, but dramatic absence of sequence similarity in the pericentromeric heterochromatin. *BMC Genomics* **16**: 374.
- Čermák, T., Curtin, S.J., Gil-Humanes, J., Čegan, R., Kono, T.J.Y., Konečná, E., Belanto, J.J., Starker, C.G., Mathre, J.W., Greenstein, R.L., and Voytas, D.F.** (2017). A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. *Plant Cell* **29**: 1196–1217.
- Clulow, S.A. and Rousselle-Bourgeois, F.** (1997). Widespread introgression of *Solanum phureja* DNA in potato (*S. tuberosum*) dihaploids. *Plant Breed.* **116**: 347–351.
- Clulow, S.A., Wilkinson, M.J., and Burch, L.R.** (1993). *Solanum phureja* genes are expressed in the leaves and tubers of aneusomatic potato dihaploids. *Euphytica* **69**: 1–6.
- Clulow, S.A., Wilkinson, M.J., Waugh, R., Baird, E., Demaine, M.J., and Powell, W.** (1991). Cytological and molecular observations on *Solanum phureja*-induced dihaploid potatoes. *Theor. Appl. Genet.* **82**: 545–551.
- Cohen, S., Houben, A., and Segal, D.** (2008). Extrachromosomal circular DNA derived from tandemly repeated genomic sequences in plants. *Plant J.*
- Comai, L.** (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**: 836–846.
- CRAN - Package MeanShift**
- Das, O.P., Poliak, E., Ward, K., and Messing, J.** (1991). A new allele of the duplicated 27kD zein locus of maize generated by homologous recombination. *Nucleic Acids Res.* **19**: 3325–3330.
- Ercolano, M.R., Carputo, D., Li, J., Monti, L., Barone, A., and Frusciante, L.** (2004). Assessment of genetic variability of haploids extracted from tetraploid ($2n = 4x = 48$) *Solanum tuberosum*. *Genome* **47**: 633–638.
- Filler Hayut, S., Melamed Bessudo, C., and Levy, A.A.** (2017). Targeted recombination between homologous chromosomes for precise breeding in tomato. *Nat. Commun.* **8**: 15605.

- Forster, B.P., Heberle-Bors, E., Kasha, K.J., and Touraev, A.** (2007a). The resurgence of haploids in higher plants. *Trends Plant Sci.* **12**: 368–375.
- Forster, B.P., Heberle-Bors, E., Kasha, K.J., and Touraev, A.** (2007b). The resurgence of haploids in higher plants. *Trends Plant Sci.* **12**: 368–375.
- Garrison, E. and Marth, G.** (2012). Haplotype-based variant detection from short-read sequencing. arXiv [q-bio.GN].
- Ghislain M., Zhang D. P., Herrera, M. R. ed** (1999). *Molecular Biology Laboratory Protocols Plant Genotyping: Training Manual* (International Potato Center: Lima, Peru).
- Gilles, L.M. et al.** (2017). Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. *EMBO J.*: e201796603.
- Hanneman, R.E. and Ruhde, R.W.** (1978). Haploid extraction in *Solanum tuberosum* group Andigena. *Am. J. Potato Res.* **55**: 259–263.
- Hardigan, M.A. et al.** (2016). Genome Reduction Uncovers a Large Dispensable Genome and Adaptive Role for Copy Number Variation in Asexually Propagated *Solanum tuberosum*. *Plant Cell* **28**: 388–405.
- Hardigan, M.A., Laimbeer, F.P.E., Newton, L., Crisovan, E., Hamilton, J.P., Vaillancourt, B., Wiegert-Rininger, K., Wood, J.C., Douches, D.S., Farré, E.M., Veilleux, R.E., and Buell, C.R.** (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. U. S. A.* **114**: E9999–E10008.
- Henry, I.M., Dilkes, B.P., Miller, E.S., Burkart-Waco, D., and Comai, L.** (2010). Phenotypic consequences of aneuploidy in *Arabidopsis thaliana*. *Genetics* **186**: 1231–1245.
- Henry, I.M., Zinkgraf, M.S., Groover, A.T., and Comai, L.** (2015). A System for Dosage-Based Functional Genomics in Poplar. *Plant Cell* **27**: 2370–2383.
- Hermesen, J.G.T. and Verdenius, J.** (1973a). Selection from *Solanum tuberosum* group Phureja of genotypes combining high-frequency haploid induction with homozygosity for embryo-spot. *Euphytica* **22**: 244–259.
- Hermesen, J. and Verdenius, J.** (1973b). Selection from *Solanum tuberosum* group Phureja of genotypes combining high-frequency haploid induction with homozygosity for embryo-spot. *Euphytica* **22**: 244–259.
- Hirsch, C.N., Hirsch, C.D., Felcher, K., Coombs, J., Zarka, D., Van Deynze, A., De Jong, W., Veilleux, R.E., Jansky, S., Bethke, P., Douches, D.S., and Buell, C.R.** (2013). Retrospective view of North American potato (*Solanum tuberosum* L.) breeding in the 20th and 21st centuries. *G3* **3**: 1003–1013.
- Iovene, M., Zhang, T., Lou, Q., Buell, C.R., and Jiang, J.** (2013). Copy number variation in potato - an asexually propagated autotetraploid species. *Plant J.* **75**: 80–89.
- Ishii, T., Karimi-Ashtiyani, R., and Houben, A.** (2016a). Haploidization via Chromosome Elimination: Means and Mechanisms. *Annu. Rev. Plant Biol.* **67**: 421–438.
- Ishii, T., Karimi-Ashtiyani, R., and Houben, A.** (2016b). Haploidization via chromosome elimination:

- means and mechanisms. *Annu. Rev. Plant Biol.* **67**: 421–438.
- Jun, G., Wing, M.K., Abecasis, G.R., and Kang, H.M.** (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* **25**: 918–925.
- Karimi-Ashtiyani, R. et al.** (2015). Point mutation impairs centromeric CENH3 loading and induces haploid plants. *Proc. Natl. Acad. Sci. U. S. A.* **112**: 11211–11216.
- Kelliher, T. et al.** (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature* **542**: 105–109.
- Kelliher, T., Starr, D., Wang, W., McCuiston, J., Zhong, H., Nuccio, M.L., and Martin, B.** (2016). Maternal Haploids Are Preferentially Induced by CENH3-tailswap Transgenic Complementation in Maize. *Front. Plant Sci.* **7**: 414.
- Kotch, G.P., Ortiz, R., and Peloquin, S.J.** (1992). Genetic analysis by use of potato haploid populations. *Genome* **35**: 103–108.
- Kuppu, S., Tan, E.H., Nguyen, H., Rodgers, A., Comai, L., Chan, S.W.L., and Britt, A.B.** (2015). Point Mutations in Centromeric Histone Induce Post-zygotic Incompatibility and Uniparental Inheritance. *PLoS Genet.* **11**: e1005494.
- Li, H.** (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio.GN].
- Liu, C. et al.** (2017). A 4-bp Insertion at ZmPLA1 Encoding a Putative Phospholipase A Generates Haploid Induction in Maize. *Mol. Plant* **10**: 520–522.
- Maheshwari, S., Tan, E.H., West, A., Franklin, F.C.H., Comai, L., and Chan, S.W.L.** (2015). Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genet.* **11**: e1004970.
- Mihovilovich, E., Aponte, M., Lindqvist-Kreuzer, H., and Bonierbale, M.** (2014). An RGA-Derived SCAR Marker Linked to PLRV Resistance from *Solanum tuberosum* ssp. *andigena*. *Plant Mol. Biol. Rep.* **32**: 117–128.
- Montelongo-Escobedo, H. and Rowe, P.R.** (1969). Haploid induction in potato: Cytological basis for the pollinator effect. *Euphytica* **18**: 116–123.
- Pâques, F. and Haber, J.E.** (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**: 349–404.
- Peloquin, S.J., Gabert, A.C., ORTIZ - Annals of botany, R., and 1996** (1996). Nature of “pollinator” effect in potato (*Solanum tuberosum* L.) haploid production. academic.oup.com.
- Pham, G.M., Braz, G.T., Conway, M., Crisovan, E., Hamilton, J.P., Laimbeer, F.P.E., Manrique-Carpintero, N., Newton, L., Douches, D.S., Jiang, J., and Others** (2019). Genome-wide inference of somatic translocation events during potato dihaploid production. *Plant Genome*.
- Pham, G.M., Newton, L., Wiegert-Rininger, K., Vaillancourt, B., Douches, D.S., and Buell, C.R.** (2017). Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. *Plant J.* **92**: 624–637.

- Pineda, O., Bonierbale, M.W., Plaisted, R.L., Brodie, B.B., and Tanksley, S.D.** (1993). Identification of RFLP markers linked to the H1 gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Genome* **36**: 152–156.
- Puchta, H.** (1999). Double-strand break-induced recombination between ectopic homologous sequences in somatic plant cells. *Genetics* **152**: 1173–1181.
- Puchta, H. and Hohn, B.** (2012). In planta somatic homologous recombination assay revisited: a successful and versatile, but delicate tool. *Plant Cell* **24**: 4324–4331.
- Qi, J., Chen, Y., Copenhagen, G.P., and Ma, H.** (2014). Detection of genomic variations and DNA polymorphisms and impact on analysis of meiotic recombination and genetic mapping. *Proc. Natl. Acad. Sci. U. S. A.*
- Ravi, M. and Chan, S.W.L.** (2010). Haploid plants produced by centromere-mediated genome elimination. *Nature* **464**: 615–618.
- Rick, C.M. and Notani, N.K.** (1961). The tolerance of extra chromosomes by primitive tomatoes. *Genetics* **46**: 1231–1235.
- Riera-Lizarazu, O., Rines, H.W., and Phillips, R.L.** (1996). Cytological and molecular characterization of oat x maize partial hybrids. *Theor. Appl. Genet.* **93**: 123–135.
- Samitsu, Y. and Hosaka, K.** (2002). Molecular marker analysis of 24- and 25-chromosome plants obtained from *Solanum tuberosum* L. subsp. *andigena* ($2n = 4x = 48$) pollinated with a *Solanum phureja* haploid inducer. *Genome* **45**: 577–583.
- Shibata, Y., Kumar, P., Layer, R., Willcox, S., Gagan, J.R., Griffith, J.D., and Dutta, A.** (2012). Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* **336**: 82–86.
- Spooner, D.M., Ghislain, M., Simon, R., Jansky, S.H., and Gavrilenko, T.** (2014). Systematics, Diversity, Genetics, and Evolution of Wild and Cultivated Potatoes. *Bot. Rev.* **80**: 283–383.
- Straadt, I.K. and Rasmussen, O.S.** (2003). AFLP analysis of *Solanum phureja* DNA introgressed into potato dihaploids. *Plant Breed.* **122**: 352–356.
- Tan, E.H., Henry, I.M., Ravi, M., Bradnam, K.R., Mandakova, T., Marimuthu, M.P., Korf, I., Lysak, M.A., Comai, L., and Chan, S.W.** (2015). Catastrophic chromosomal restructuring during genome elimination in plants. *Elife* **4**.
- Tomson, B.N., D'Amours, D., Adamson, B.S., Aragon, L., and Amon, A.** (2006). Ribosomal DNA transcription-dependent processes interfere with chromosome segregation. *Mol. Cell. Biol.* **26**: 6239–6247.
- Velásquez, A.C., Mihovilovich, E., and Bonierbale, M.** (2007). Genetic characterization and mapping of major gene resistance to potato leafroll virus in *Solanum tuberosum* ssp. *andigena*. *Theor. Appl. Genet.* **114**: 1051–1058.
- Wagenvoort, M. and Lange, W.** (1975). The production of aneuidihaploids in *Solanum tuberosum* L. group *Tuberosum* (the common potato). *Euphytica* **24**: 731–741.
- Wangenheim, K.H., Peloquin, S.J., and Hougas, R.W.** (1960). Embryological investigations on the

formation of haploids in the potato (*Solanum tuberosum*). *Molecular and General*.

Watanabe, K.N. and Orrillo, M. (1993). An alternative pretreatment method for mitotic chromosome observation in potatoes. *Am. Potato J.* **70**: 543–548.

Waugh, R., Baird, E., and Powell, W. (1992). The use of RAPD markers for the detection of gene introgression in potato. *Plant Cell Rep.* **11**: 466–469.

Wijnker, E. et al. (2013). The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. *Elife* **2**: e01426.

Wilkinson, M.J., Bennett, S.T., Clulow, S.A., Allainguillaume, J., Harding, K., and Bennett, M.D. (1995). Evidence for somatic translocation during potato dihaploid induction. *Heredity* **74 (Pt 2)**: 146–151.

Zhao, X., Xu, X., Xie, H., Chen, S., and Jin, W. (2013). Fertilization and uniparental chromosome elimination during crosses with maize haploid inducers. *Plant Physiol.* **163**: 721–731.

Zinkgraf, M., Haiby, K., Lieberman, M.C., Comai, L., Henry, I.M., and Groover, A. (2017). Creation and genomic analysis of irradiation hybrids in *Populus*. *Current Protocols in Plant Biology* **1**: 431–450.

Figures

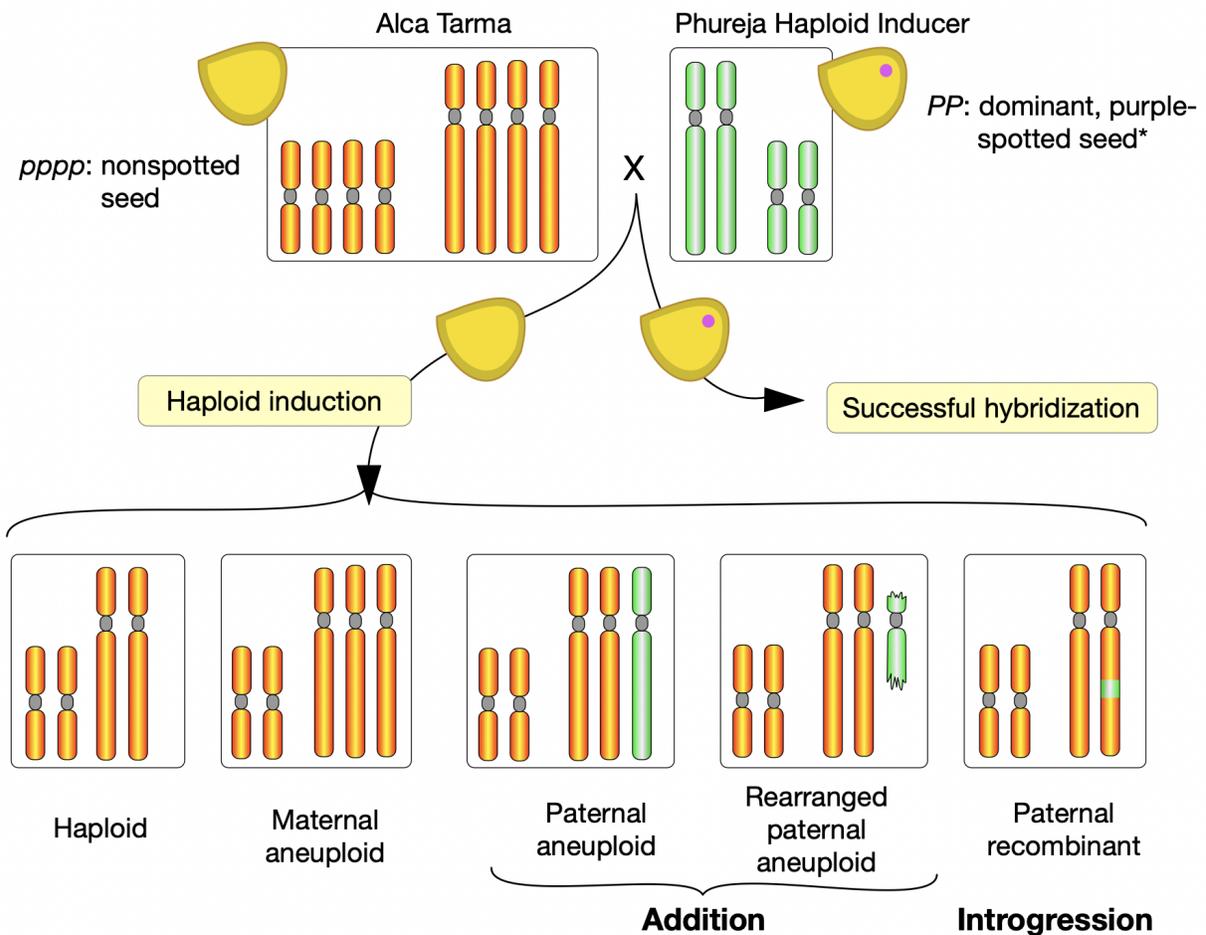


Figure 2.1. Production of the LOP population and expected types. Haploid inducers IVP101 or PL4, both diploids, were used to pollinate tetraploid cultivar Alca Tarma. For simplicity, the potato genome is represented with two chromosome types. The haploid inducer is homozygous for the dominant seed purple spot marker. Normal fertilization and development resulting in hybrids with spotted seed, defined as successful hybridization, results in biparental triploids or tetraploids, depending on the ploidy of the male gamete. Maternal dihaploids are expected among plants germinated from nonspotted seeds *(Hermsen and Verdenius, 1973b). Plants displaying more than 8 stomatal chloroplasts (an indication of increased nuclear content) or unusual phenotypes potentially consistent with aneuploidy, were discarded (Velásquez et al., 2007; Mihovilovich et al., 2014). Genetic haploid inducers can act through either parthenogenesis (Forster et al., 2007b) (development of an unfertilized egg) or genome elimination (Ishii et al., 2016b) (rejection of the haploid inducer genome). Addition or introgression of residual haploid inducer DNA indicates the second mode of action.

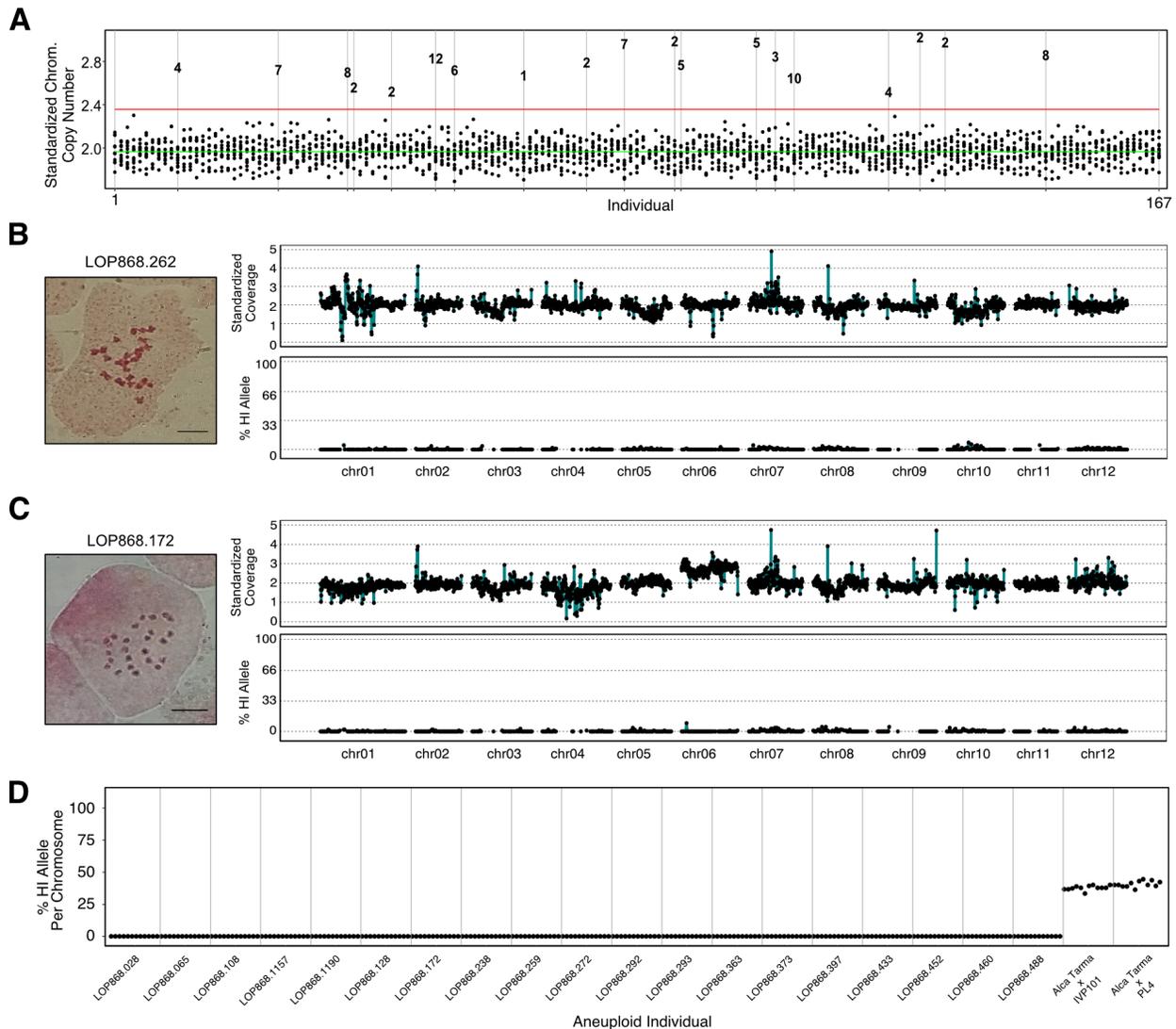


Figure 2.2. Aneuploidy detection in 167 primary dihaploids by genome sequencing. (A)

Chromosome copy number for each individual in the population. Each individual of the population is displayed along the X-axis, with the stack of dots at each X coordinate corresponding to standardized copy number for each chromosome type, with a value of 2.0 representing the expected diploid state. The green line corresponds to the mean chromosome copy number among the population and the red line indicates 3 standard deviations greater than the mean. Outliers in this distribution correspond to the

affected chromosome in each trisomic and are numbered according to the chromosome present in excess. **(B)** Cytogenetic and *in silico* karyotype of a representative euploid dihaploid LOP868.262. Left: Root tip somatic metaphase karyotype. Top right: copy number plot; individual data points represent read depth in non-overlapping 250kb bins standardized to tetraploid Alca Tarma counts such that the expected diploid state corresponds to copy number 2.0. Bottom right: haploid inducer SNP allele plot; black points correspond to the fraction of base calls supporting haploid inducer alleles at all informative sites in non-overlapping 1Mb bins. **(C)** Cytogenetic and *in silico* karyotype of chromosome 6 trisomic LOP868.172, illustrating that trisomy of chromosome 6 was not derived from either IVP101 or PL4. See Supplemental Figure S2.1 for dosage plots of the remaining trisomics. **(D)** SNP plot showing the percentage of haploid inducer allele present in the trisomics identified in panel A. For each individual, the 12 points correspond to the 12 chromosomes displayed in order (i.e., the first dot is chromosome 1, the second chromosome 2, etc). For each individual, points near 0% for the affected chromosome and all others indicate a maternal trisomy. Observed % HI allele values from two representative simulated hybrid controls, one Alca Tarma x IVP101, the other Alca Tarma x PL4, are shown on the right.

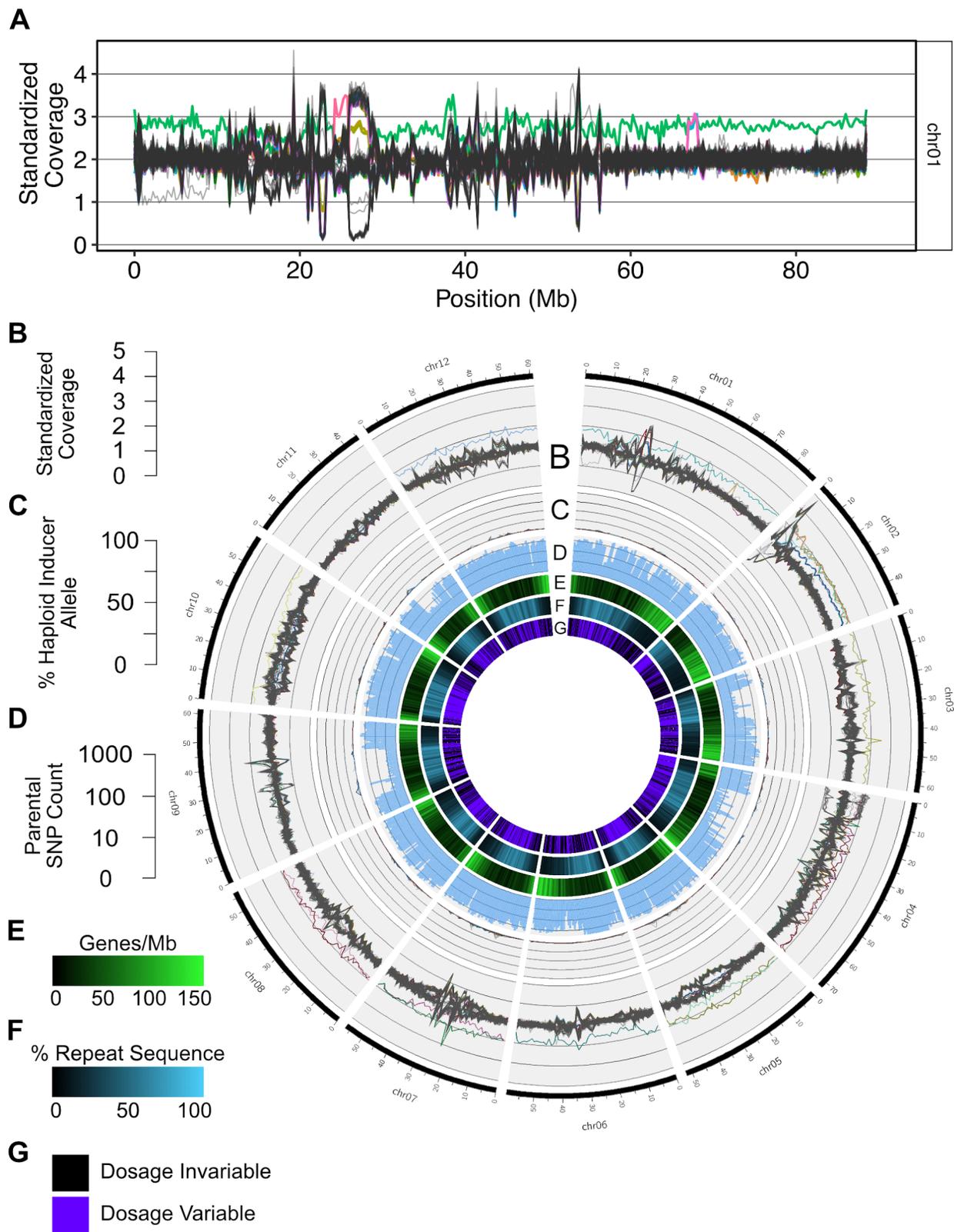


Figure 2.3. Chromosomal distributions of dosage variation and parent-informative markers in the LOP dihaploid population. **A)** Read depth standardized to tetraploid Alca Tarma in non-overlapping 250kb bins of chromosome 1. Each individual in the dihaploid population is represented by a single line. Dihaploids that are trisomic for any chromosome are represented by colored lines. For example, chromosome 1 trisomic LOP868.238 is displayed as a green line. The plot also displays unique segmental dosage variants, one starting at 24 Mb and involving a trisomic of 8 (pink line), a second one at 67 Mb and involving two trisomics of 2 (overlapping magenta and blue lines), and a terminal deletion of the short arm in an otherwise euploid line (grey line). **B-F)** Circos plots for the 12 chromosomes. For each chromosome, 167 lines are shown. Aneuploids of any chromosome are uniquely colored and all other individuals are colored gray. **B)** Standardized read depth plots of all chromosomes smoothed to 1Mb bins. **C)** Percent haploid inducer allele at all parent-informative marker loci in non-overlapping 1Mb bins. Refer to Fig. S2.3 for expected haploid inducer allele percentages derived from simulated hybrid analyses. **D)** Log₁₀-scaled counts of parent-informative markers in non-overlapping 1Mb bins. **E)** Genes per 1Mb sliding window, 200kb step. **F)** Percent repeat sequence per 1Mb sliding window, 200kb step.

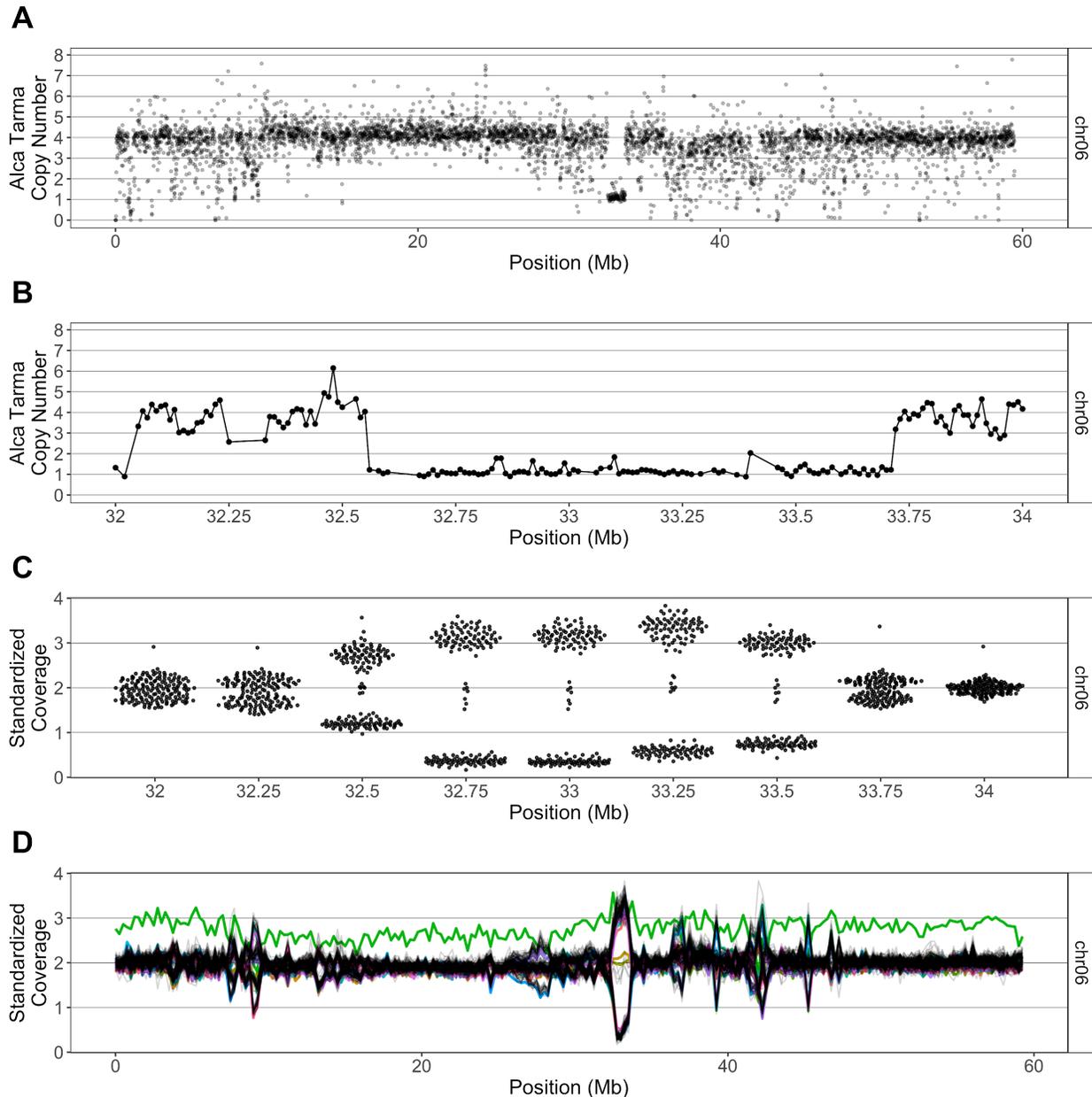
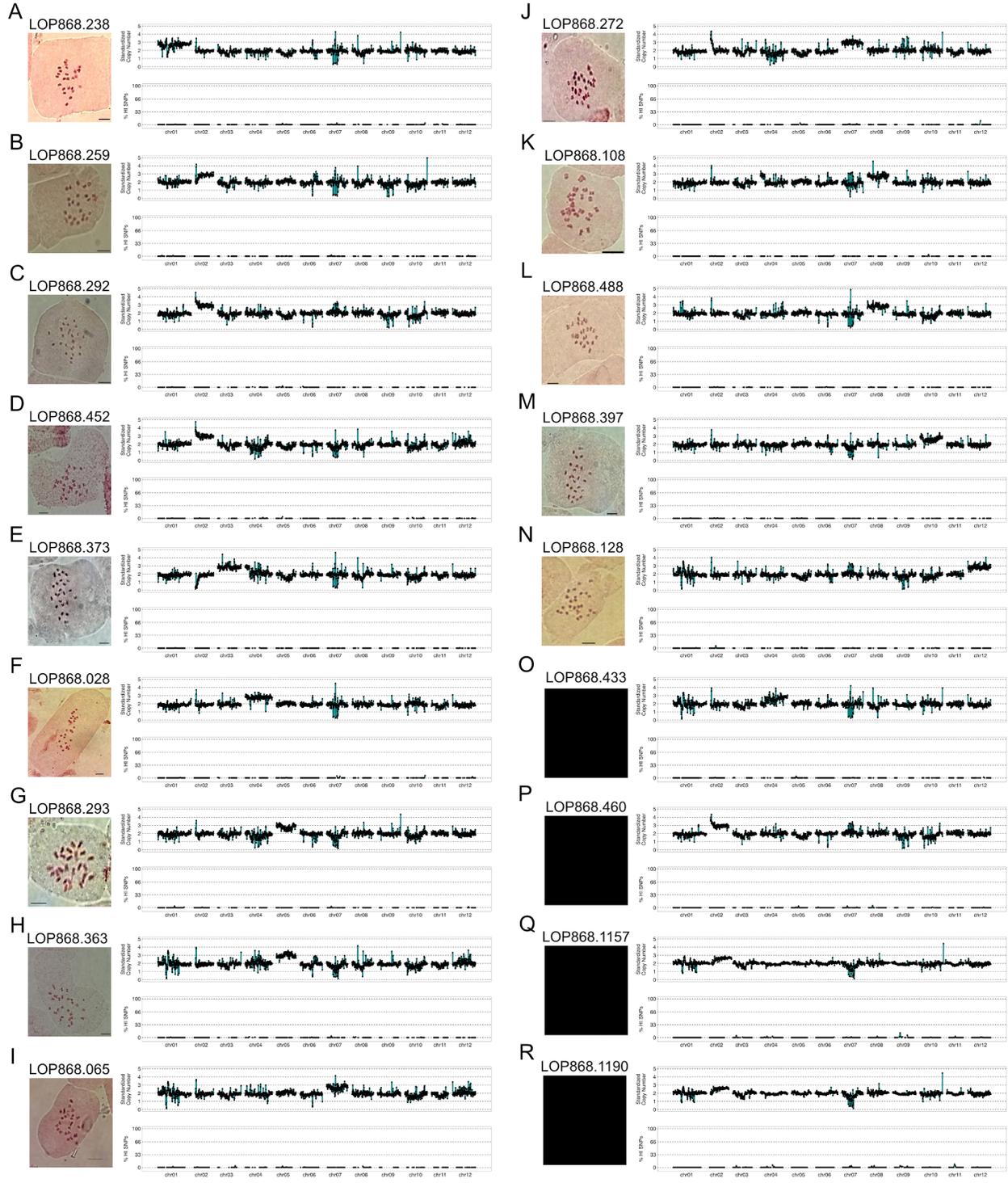


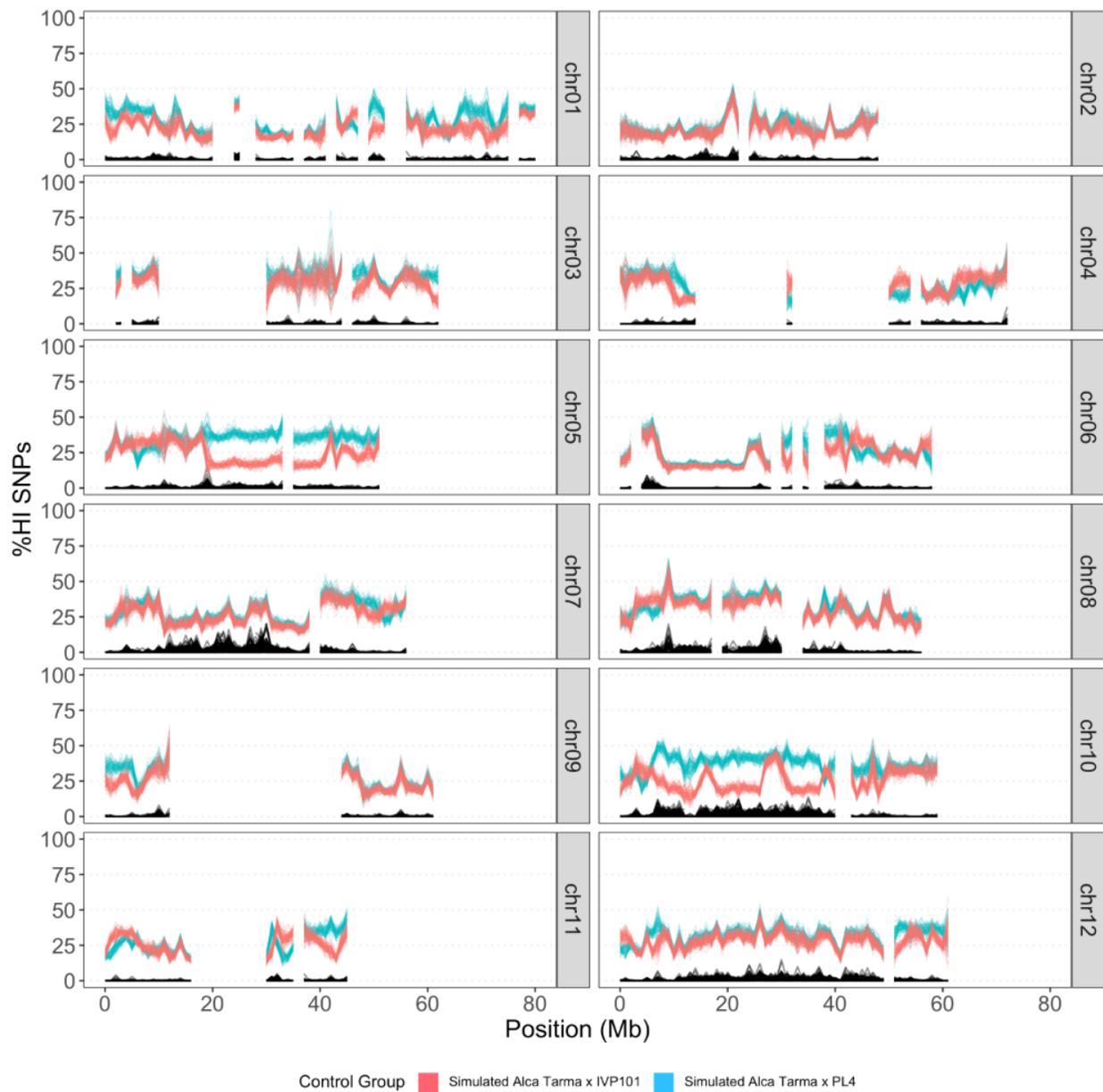
Figure 2.4. Example of segregating CNV among the LOP dihaploid population. A) GC-normalized median read depth of non-overlapping 10kb bins for tetraploid Alca Tarma (see Methods). Chromosome 6 is shown. Each dot is plotted with partial transparency to emphasize overplotting. **B)** GC-normalized median read depth, zoomed in on 1Mb region of chromosome 6. **C)** Swarm plots of standardized coverage values in bins affected by 1Mb deletion. The population segregates the high and low dosage states in an approximate 1:1 ratio, consistent with random chromosome segregation of a deletion in triplex allele dosage. **D)** Population standardized values for chromosome 6. Each dihaploid is displayed as a single contiguous line, Green line: chromosome 6 trisomic. The region corresponding to the large deletion shown in panel A exhibits segregating dosage variation in the dihaploid population.

Supplemental material

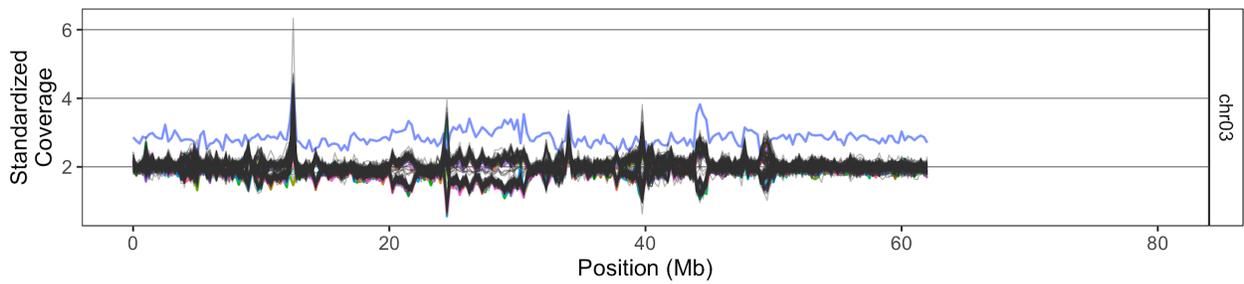
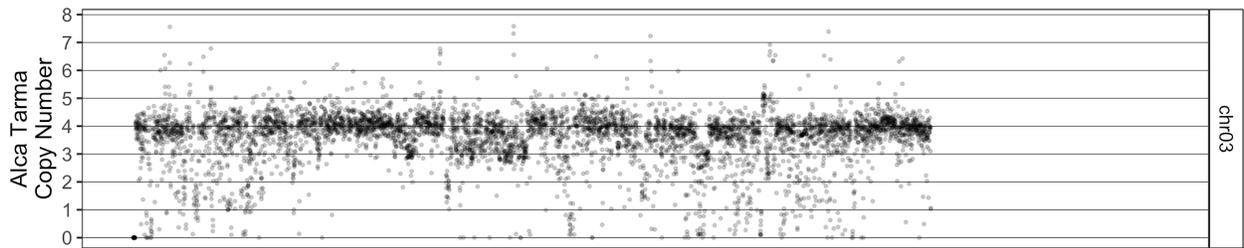
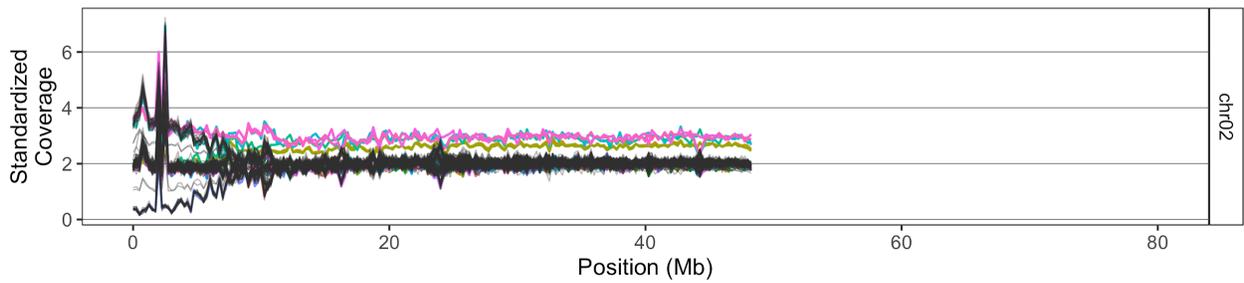
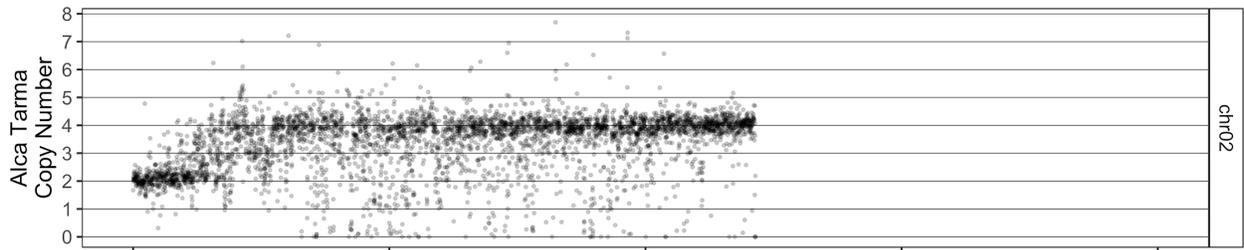
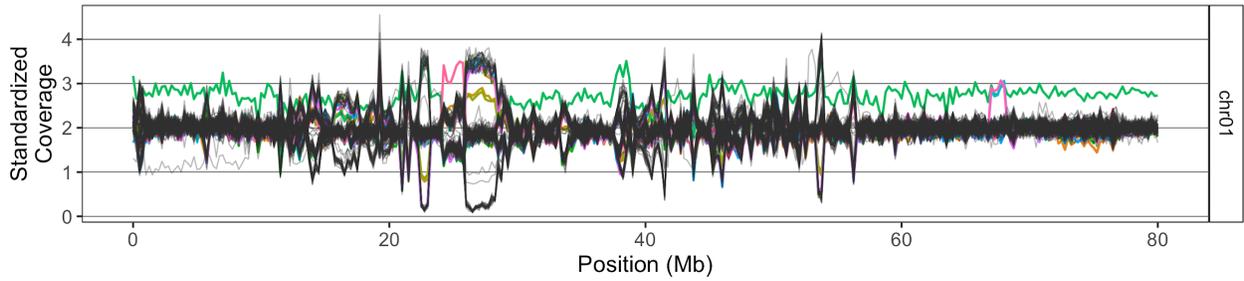
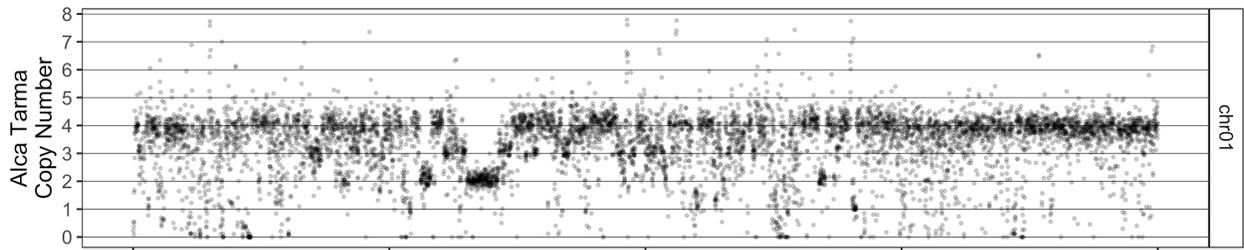
Supplemental Figures

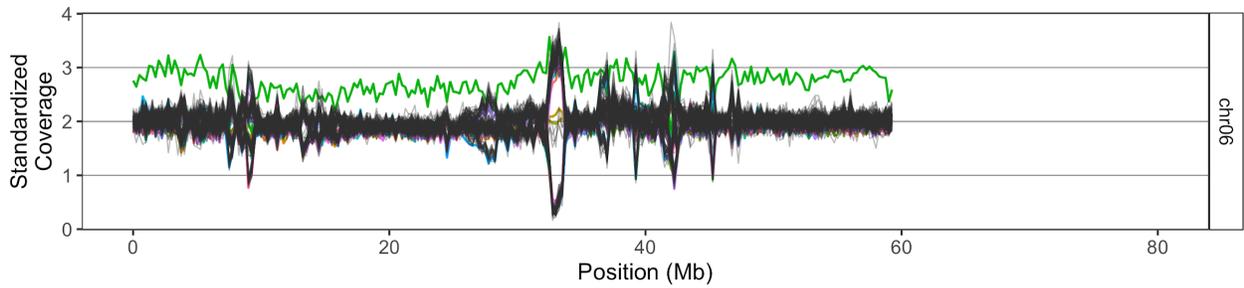
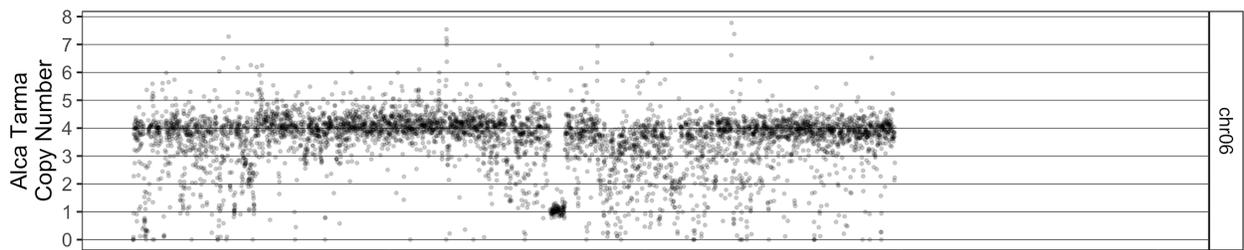
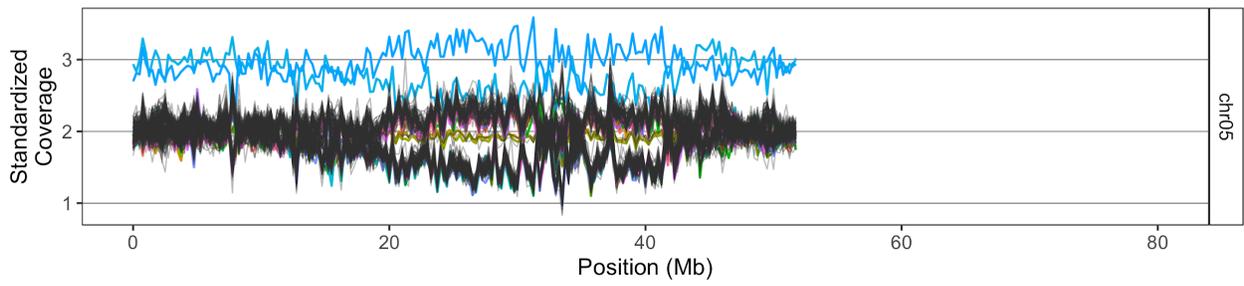
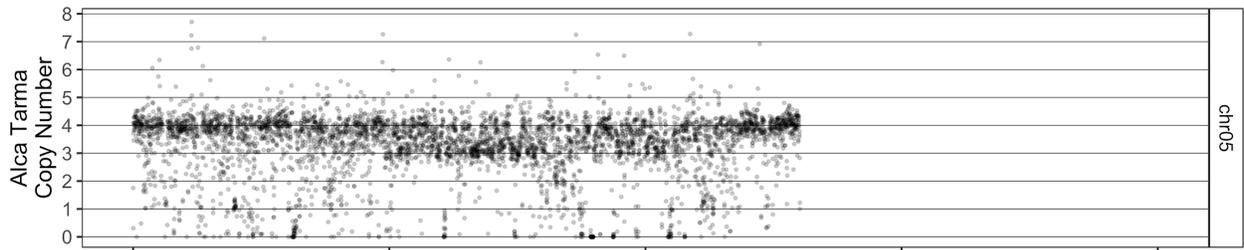
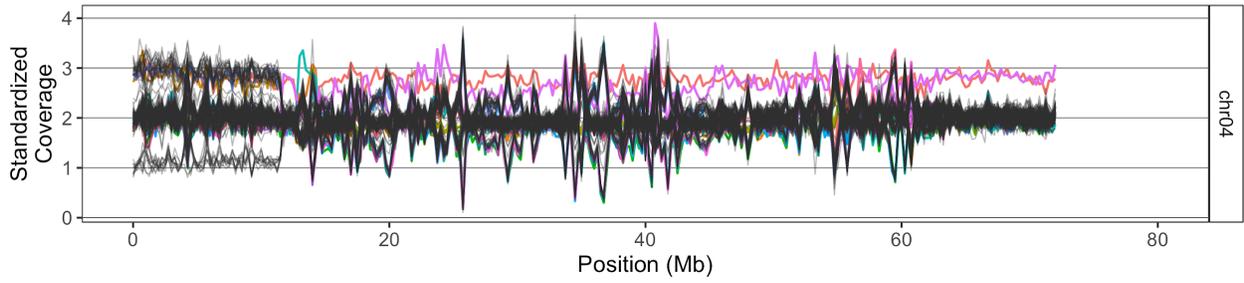
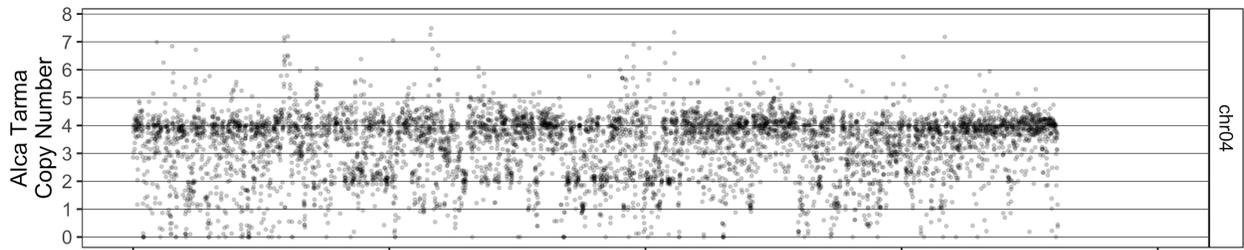


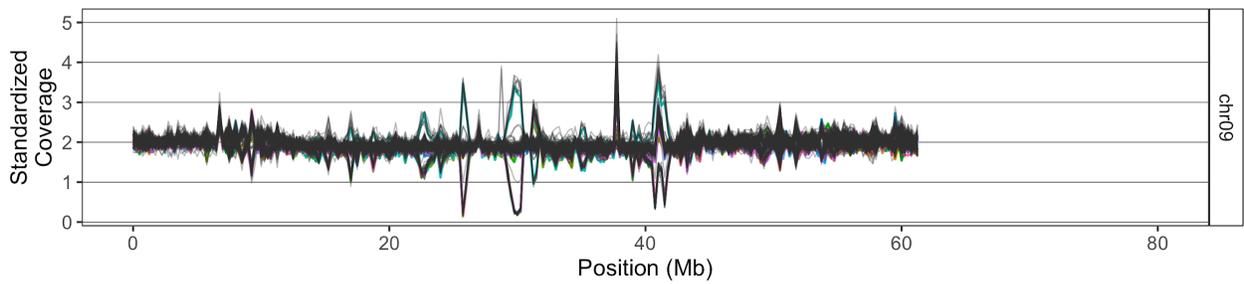
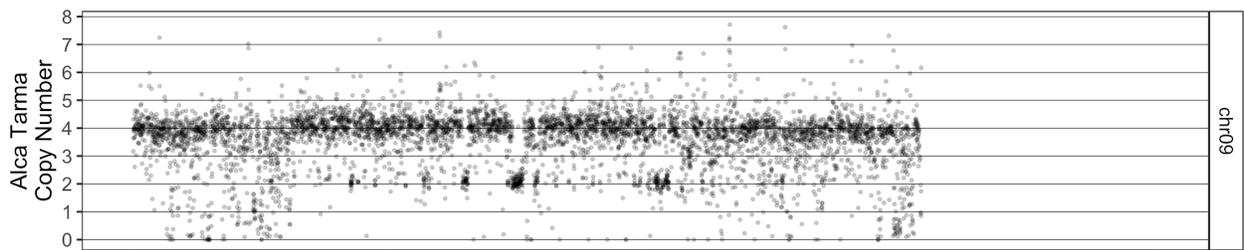
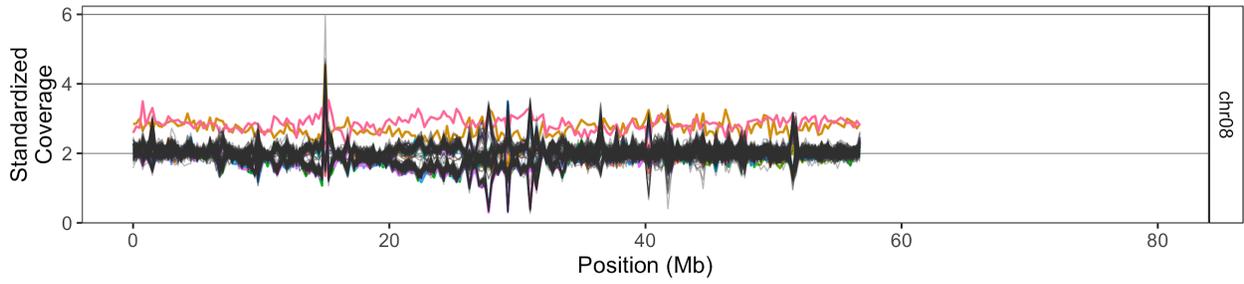
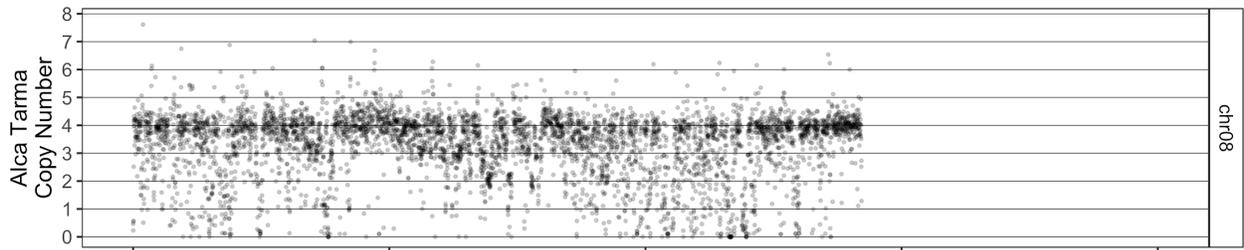
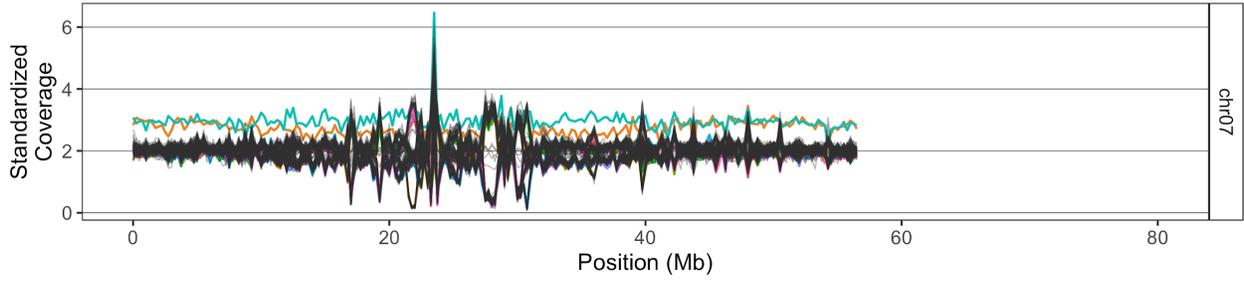
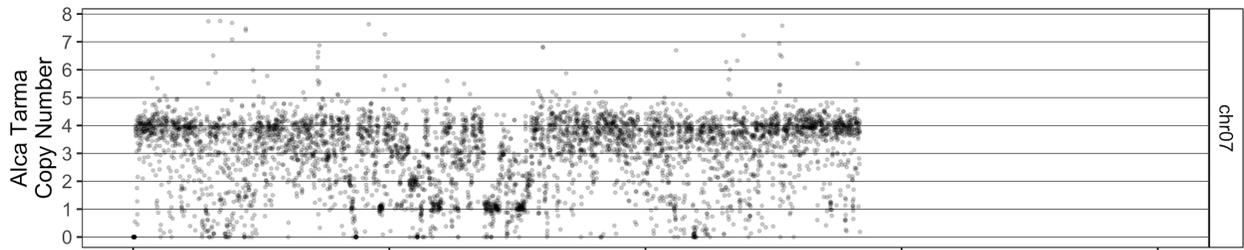
Supplemental Figure S2.1. Aneuploidy detection. (A-R) Dosage plots and somatic chromosome spreads of putative trisomics. Black boxes in places of karyotypes indicate that a putative trisomic was not available for chromosome counting. Bars: 5 μ m.

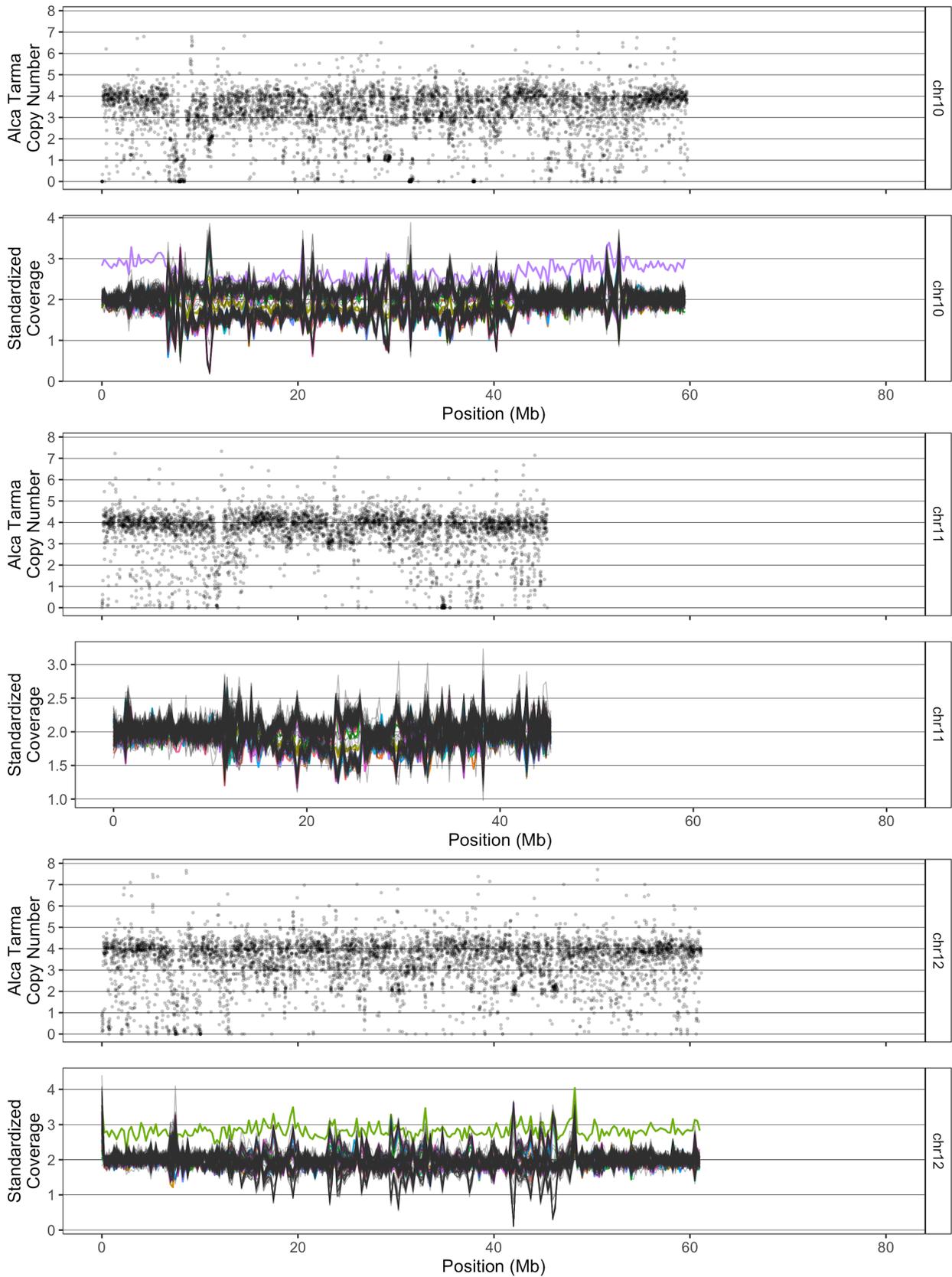


Supplemental Figure S2.2. Power analysis for HI SNP measured in dihaploid by low pass sequencing. Each datapoint illustrates the fraction of haploid inducer allele among reads aligning at all parent-informative loci in a 1 Mb bin. Black: observed HI SNP incidence in the dihaploid population. Red (IVP101) and blue (PL4): resampled HI SNP modeling for each bin the distribution of expected measurements from a hypothetical *Hhh* hybrid (*H*: HI allele, *h*: Alca Tarma allele).





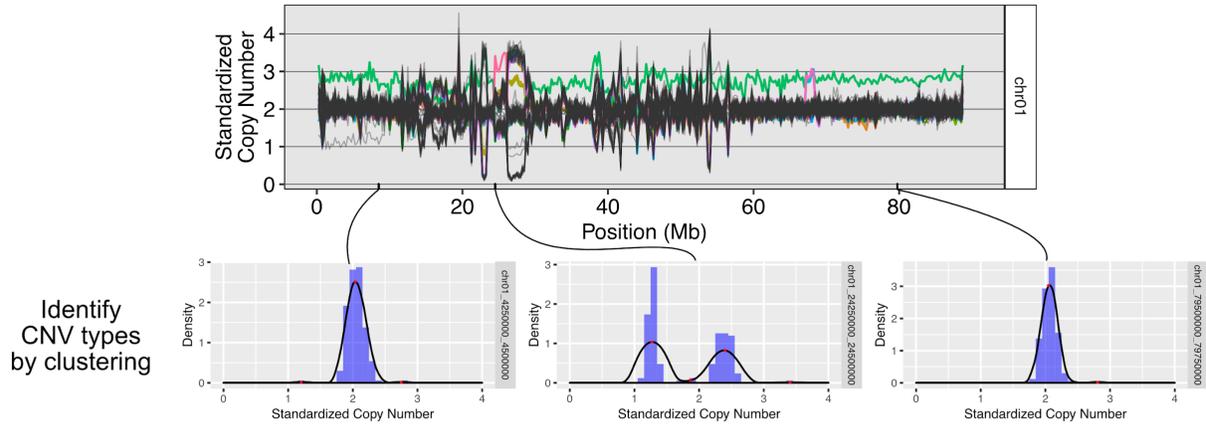




Supplemental Figure S2.3. Expanded view of dosage variation in dihaploid population. For each

pair of panels, the upper panel indicates estimated copy number in tetraploid Alca Tarma in non-overlapping 10kb bins for one chromosome. After excluding bins with $\geq 30\%$ N content and using only reads with mapping quality $\geq Q20$ in each window, median read depth was divided by bin GC content. Copy number 4 is scaled to the mode of GC-normalized read depth values across all bins. Each lower panel depicts relative coverage values in non-overlapping 250kb bins, with each line corresponding to the values for one dihaploid. Lines corresponding to each trisomic dihaploid are given the same color across all 12 chromosomes. To provide optimal resolution for each chromosome, the Y-axes of each lower panel are scaled independently.

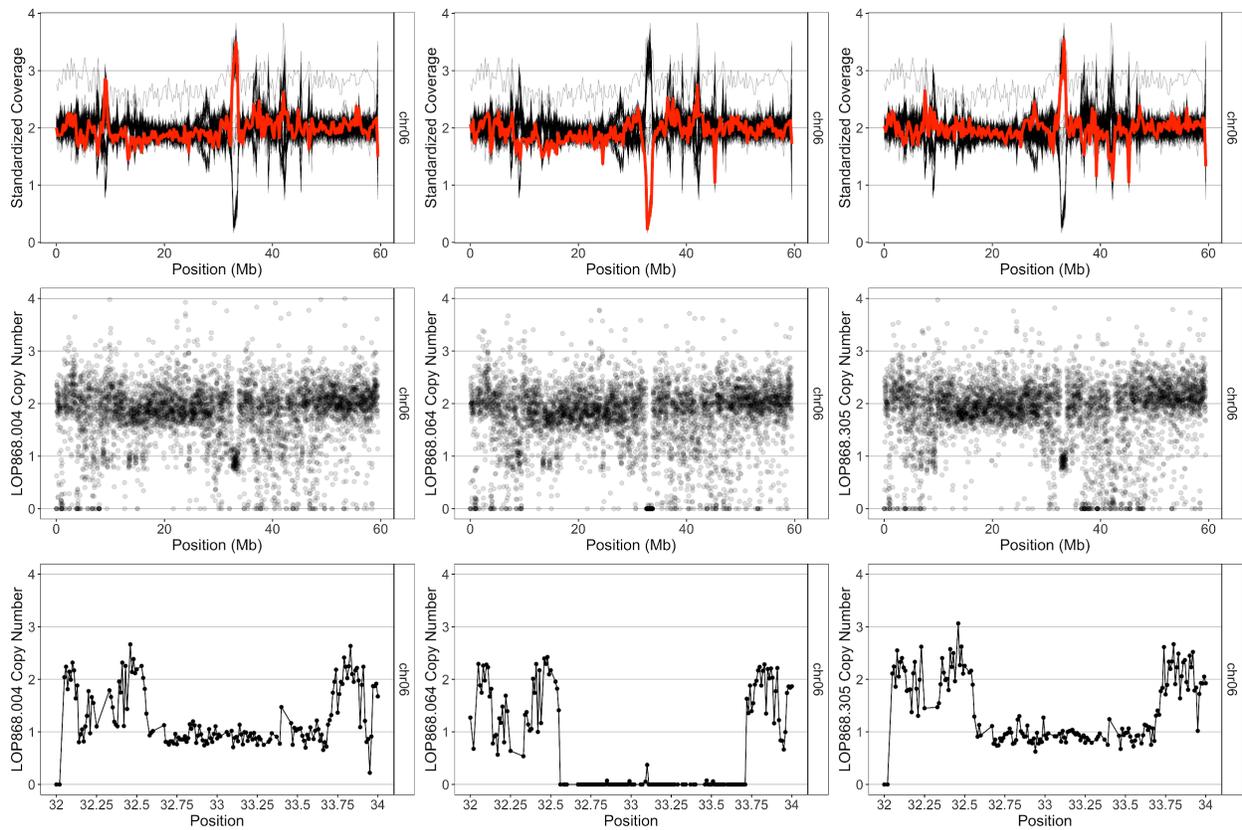
Overlay 167 standardized coverage plots of chromosome 1



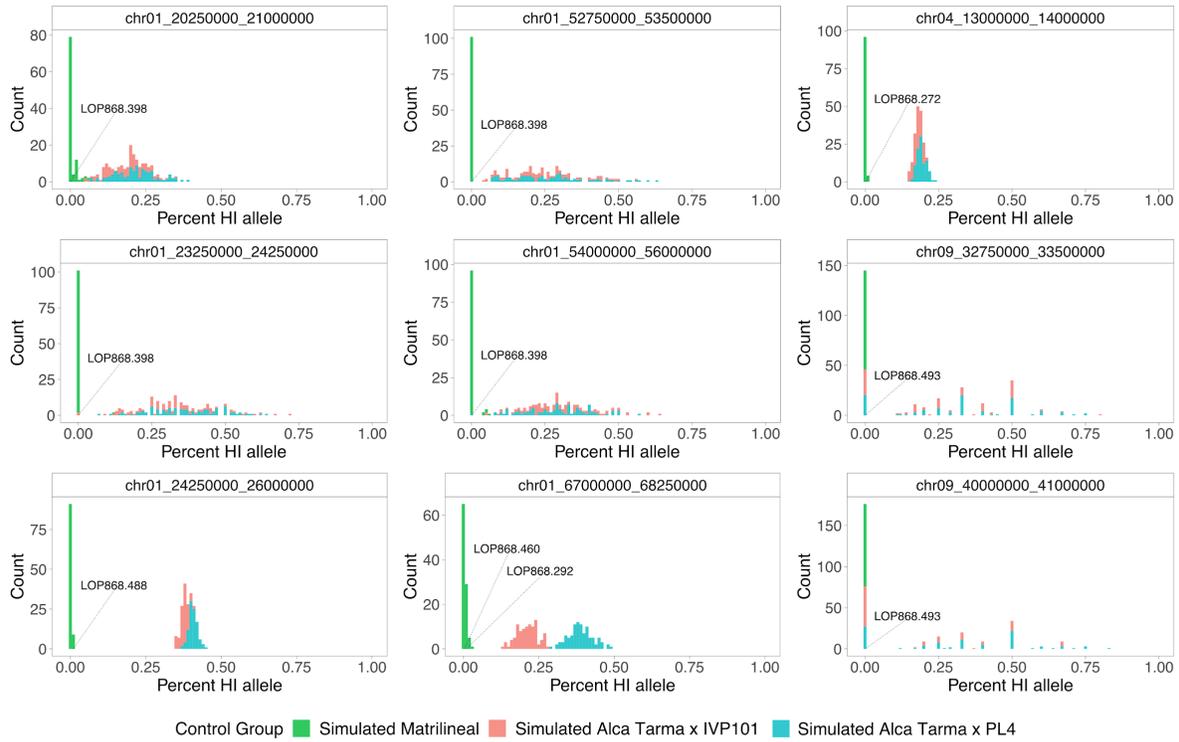
Identify CNV types by clustering

	chr01_4250000_4500000		chr01_24250000_24500000		chr01_79500000_79750000	
Dihaploid	Standardized CN	Cluster	Standardized CN	Cluster	Standardized CN	Cluster
LOP868.004	2.019677	2	1.319201	1	1.871437	1
LOP868.005	2.140322	2	2.356690	2	2.035361	1
LOP868.238	2.753821	3	1.207886	1	2.810611	2
LOP868.305	2.225104	2	2.325338	2	2.028786	1
LOP868.398	1.206732	1	1.446455	1	2.121506	1
LOP868.488	1.934082	2	3.410949	1	2.071303	1
....

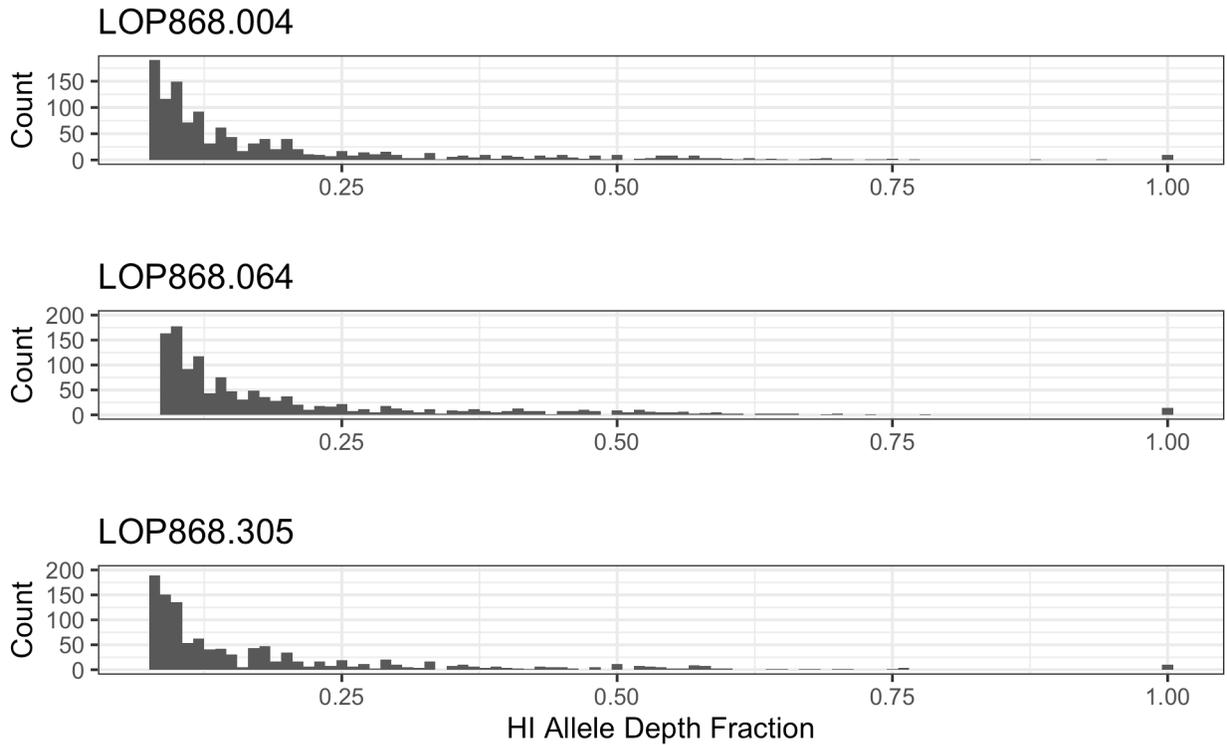
Supplemental Figure S2.4. Assignment of CNV type to each dihaploid line. For each bin, dosage states are defined as distinct clusters of standardized coverage values. Bins displaying two or more clusters are consistent with underlying structural polymorphism among Alca Tarma haplotypes. Outlier clusters were defined as those with fewer than three constituent members.



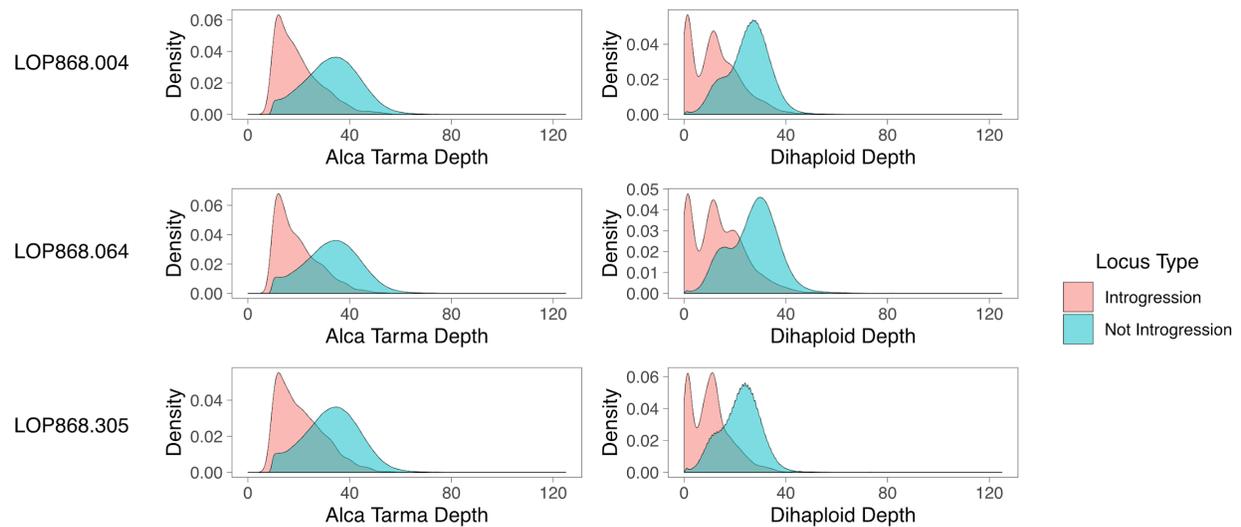
Supplemental Figure S2.5. CNV inference from dosage-variable states. Data from dihaploids LOP868.004, LOP868.064, and LOP868.305 are shown from left to right. Top panels: standardized coverage from low coverage data of each dihaploid in red overlaid against the standardized coverage value from the population in black. Middle panels: median read depth in non-overlapping 10kb windows, normalized by bin GC content. Chromosome 6 is shown. Lower panels: median read depth in non-overlapping 10kb windows, normalized by bin GC content. Shown is a region of chromosome 6 affected by a polymorphic ~1Mb deletion.



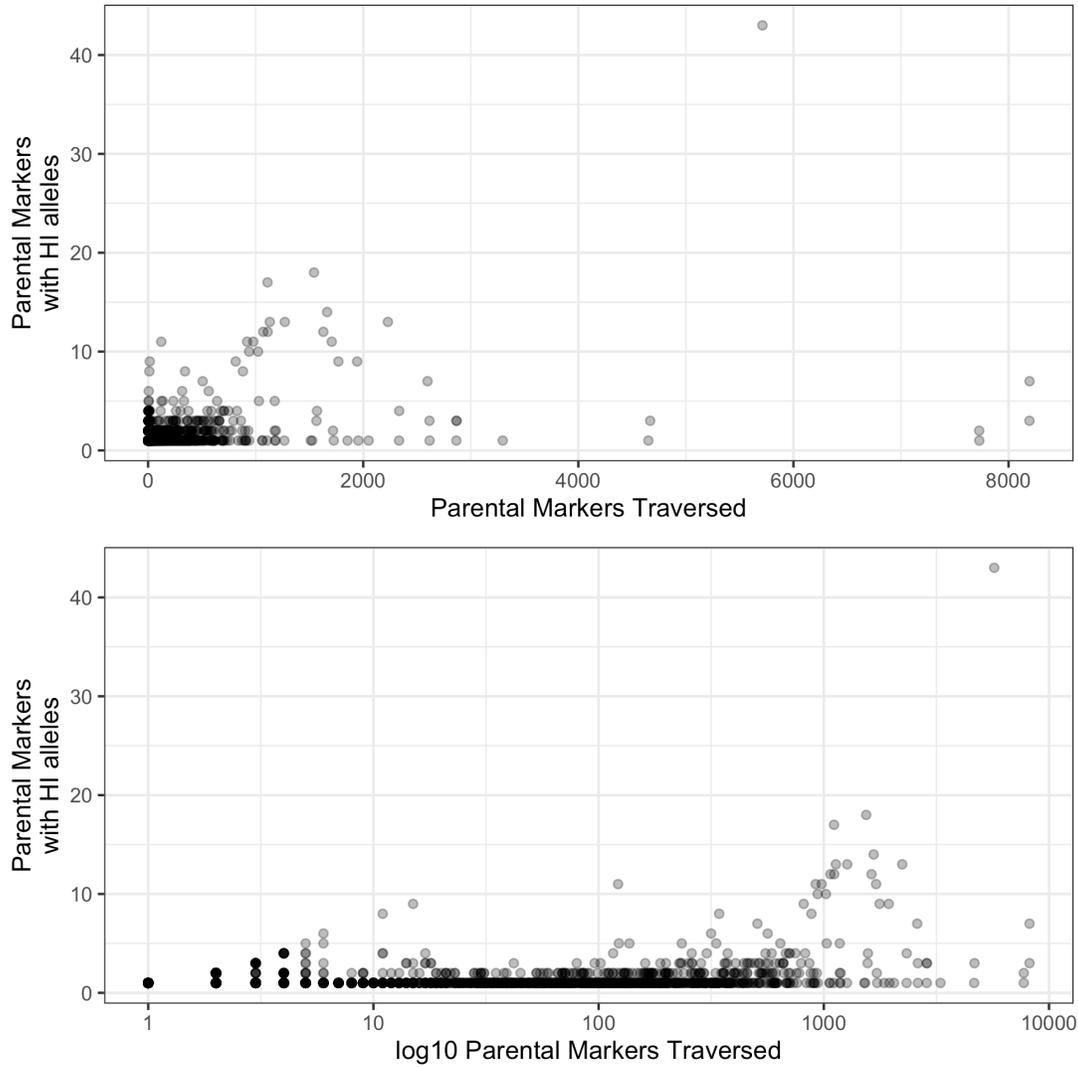
Supplemental Figure S2.6. Power analysis for HI SNP measured by low pass sequencing for identified rare structural variants. Each panel represents a histogram of observed HI SNP allele obtained from 100 simulated matrilineal samples (green), 100 simulated triploid Alca Tarma x IVP101 hybrids (red), or 100 simulated Alca Tarma x PL4 hybrids (blue) at a locus corresponding to an identified rare dosage variant in the dihaploid population. Bins were called resolvable if complete separation was observed between the simulated matrilineal group and both simulated hybrid control groups. Observed %HI for bin chr09:40.25-41Mb, which were absent for all simulated control groups due to insufficient marker density, are not shown.



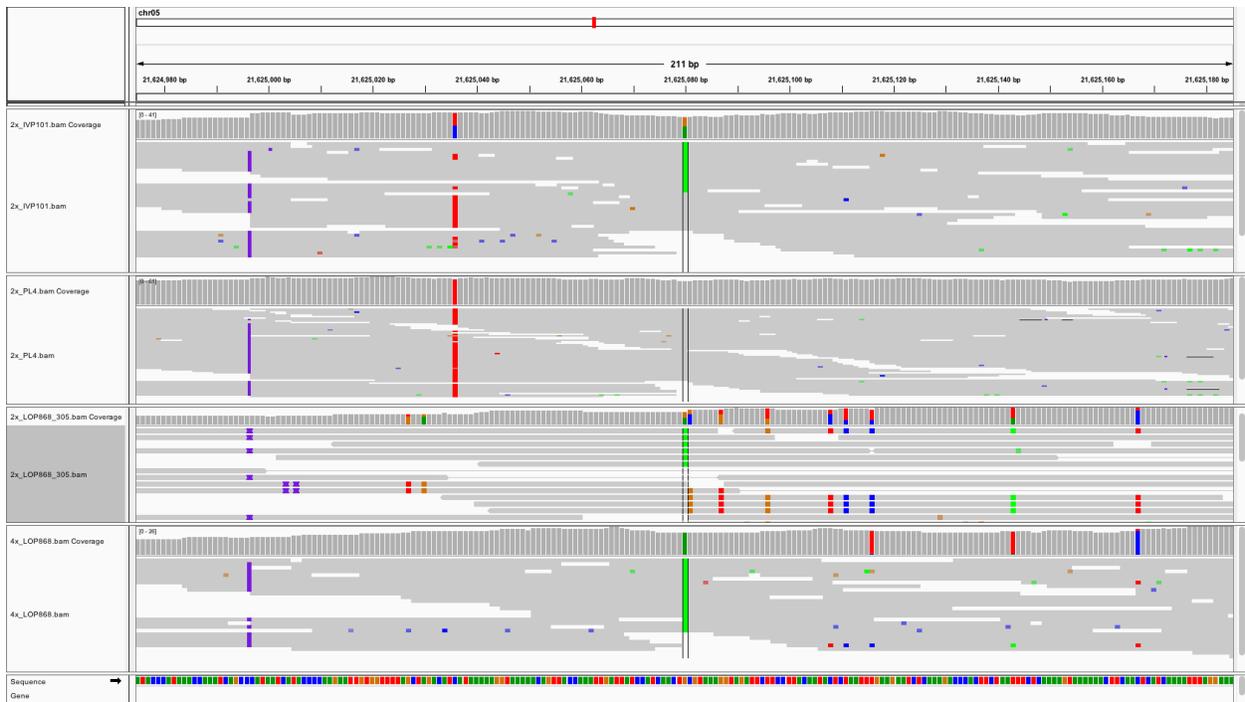
Supplemental Figure S2.7. Histograms of percent of haploid inducer alleles at putative introgression loci. At each putative introgression locus, reads matching a haploid inducer-specific allele were counted and divided by the total read depth at that locus. The resulting fractions at each putative introgression locus are then tallied, binned in increments of 0.01, and plotted.



Supplemental Figure S2.8. Read depth comparison between putative introgression-positive and introgression-negative sites in three dihaploids. Top panels: Read depth distributions at introgression and non-introgression SNP loci identified in LOP868.004. Shown on the left is the read depth distribution for Alca Tarma; on the right, dihaploid LOP868.004. Middle panels: Read depth distributions at introgression and non-introgression SNP loci identified in LOP868.064. Shown on the left is the read depth distribution for Alca Tarma; on the right, read depth of dihaploid LOP868.064. Lower panels: Read depth distributions at introgression and non-introgression SNP loci identified in LOP868.305. Shown on the left is the read depth distribution for Alca Tarma; on the right, read depth of dihaploid LOP868.305.



Supplemental Fig. S2.9. Assessment of marker conversion rate for each introgression event in three *Alca Tarma* dihaploids. Each data point illustrates an introgression event in one of the three high-coverage dihaploids. The X-axis corresponds to the number of parental SNP traversed in each event. Shown on the Y-axis is the number of parental SNP traversing a putative introgression event that exhibited haploid inducer alleles. The upper and lower panels show the same data; the lower is log10 scaled on the X-axis to illustrate marker conversion rate at low and intermediate numbers of traversed markers.



Supplemental Figure S2.10. Genome browser screenshot of representative false positive SNP locus. Physical phasing of non-reference nucleotides indicates read alignments that support presence of a haploid inducer allele in dihaploid LOP868.305 appear artifactual, as they do not match a local haplotype present in either haploid inducer. Of 43 examined loci on chromosome 5, all exhibited this pattern. The complete set of screenshots is provided as Supplemental Data Set S2.4.

Supplemental Tables

Supplemental Table S2.1. Summary of sequencing data.

Supplemental Table S2.2. Rare dosage variants and associated haploid inducer SNP profiles.

Supplemental Data Sets

Supplemental Data Set S2.1. List of homozygous parental SNP used for chromosome-wide dosage analysis.

Supplemental Data Set S2.2. List of parental SNP (homozygous and heterozygous haploid inducer genotypes included) used for segmental dosage analysis.

Supplemental Data Set S2.3. Filtered genotype calls from high coverage data of LOP-868, IVP101, PL4, and three selected dihaploids.

Supplemental Data Set S2.4. Genome browser screenshots of loci corresponding to putative introgression event with the highest number of converted loci.

All supplemental figures and datasets are available from Figshare at

https://figshare.com/projects/Genomic_Outcomes_of_Haploid_Induction_Crosses_in_Potato_Solanum_tuberosum_L_/70793

Chapter 3

Rare instances of haploid inducer DNA in potato dihaploids and ploidy-dependent genome instability

[Published in: The Plant Cell, koab100]

Kirk R Amundson¹, Benny Ordoñez^{1,2,6}, Monica Santayana², Mwaura Livingstone Nganga¹, Isabelle M Henry¹, Merideth Bonierbale^{2,4}, Awais Khan^{2,5}, Ek Han Tan³, and Luca Comai^{1*}

¹ Plant Biology and Genome Center University of California, Davis, 1 Shields Avenue, Davis, CA, 95616, USA.

² International Potato Center (CIP), P.O. Box 1558, Lima 15024, Peru.

³ School of Biology and Ecology, University of Maine, Orono, ME 04469, USA.

⁴ Current Address: Duquesa Business Centre, P.O. Box 157, Manilva, Malaga, Spain 29692.

⁵ Current Address: Plant Pathology and Plant-Microbe Biology Section, Cornell University, 14456 Geneva, NY, USA.

⁶ Integrative Genetics and Genomics Graduate Group, University of California, Davis, CA 95616, USA.

*Corresponding author: lcomai@ucdavis.edu

Short title: Paternal contribution to potato haploids

The author responsible for distribution of materials integral to the findings presented in the article in accordance with the policy described in the Instructions for Authors is: Luca Comai (lcomai@ucdavis.edu)

Abstract

In cultivated tetraploid potato, reduction to diploidy (dihaploidy) allows hybridization to diploid germplasm, introgression breeding, and may facilitate the production of inbreds. Pollination with haploid inducers yields maternal dihaploids, as well as triploid and tetraploid hybrids. It is not known if dihaploids result from parthenogenesis, entailing development of embryos from unfertilized eggs, or genome elimination, entailing missegregation and loss of paternal chromosomes. A sign of genome elimination is the occasional persistence of haploid inducer DNA in some of the dihaploids. We characterized the genomes of 919 putative dihaploids and 134 hybrids produced by pollinating tetraploid clones with three haploid inducers, IVP35, IVP101, and PL4. Whole-chromosome or segmental aneuploidy was observed in 76 dihaploids with karyotypes ranging from $2n=2x-1=23$ to $2n=2x+3=27$. Of 74 aneuploids with additional chromosomes, 66 contained chromosomes from the non-inducer parent and 8 showed chromosomes from the inducer parent. Chromosomal breaks commonly affected the paternal genome in the dihaploid and tetraploid progeny, but not in the triploid progeny, correlating instability to sperm ploidy and to haploid induction. Residual haploid inducer DNA is consistent with genome elimination as the mechanism of haploid induction.

Introduction

Cultivated potato (*Solanum tuberosum* L.) is predominantly autotetraploid ($2n=4x=48$), vegetatively propagated, highly heterozygous, and can be severely affected by inbreeding depression. These attributes make potato improvement through conventional breeding slow and difficult, and have renewed efforts to reinvent potato as a diploid and inbred-based crop based on true seed, in order to keep pace with a rapidly changing market and climate (Jansky et al., 2016).

The first step is to capture useful genetic diversity of elite tetraploid cultivars at the diploid level (Jansky et al., 2016; Lindhout et al., 2011). This can be routinely achieved through pollination of a tetraploid of interest with select clones that act as a Haploid Inducer (HI). In a 4x by 2x HI cross, a fraction of the progeny are $2n=2x=24$ primary dihaploids lacking chromosomes from the HI parent. Several diploid accessions of Andigenum group potato (formerly *S. phureja* (Spooner et al., 2014)) were demonstrated to act as efficient HIs over half a century ago (Gabert, 1963), and subsequent breeding efforts were successful in obtaining more efficient HIs, as well as incorporating a dominant marker that aids in distinguishing dihaploids from hybrids, which are usually discarded (Hermsen and Verdenius, 1973; Hutten et al., 1993). In practice, most hybrids from 4x by 2x crosses of potato are tetraploid rather than triploid, presumably because of a triploid block (Marks, 1966; Hanneman and Ruhde, 1978; Jackson et al., 1978; Hanneman and Peloquin, 1968).

Relatively little is known about the molecular basis of haploid induction in potato, but cytological evidence provides some clues. In a 4x WT by 2x HI cross, dihaploids originate from seeds with hexaploid (6x) endosperm, which is the expected outcome of a 4x central cell fertilization by a 2x sperm (Wangenheim et al., 1960). In different HI clones, 30-40% of pollen fails to complete the second mitosis, resulting in a single, larger restitution sperm, or occasionally, two sperms that potentially have unbalanced chromosome sets (Montelongo-Escobedo and Rowe, 1969; Montezuma-de-Carvalho, 1967). Furthermore, colchicine-treated pollen, but not untreated pollen of *S. tarijense*, develops restitution sperm and induces potato dihaploids, and colchicine treatment of HI pollen further increases the haploid induction rate (Montelongo-Escobedo and Rowe, 1969). Unreduced 2x sperm are also produced from

restitution of the first or second meiotic divisions, but this increased rate of meiotic restitution is not associated with increased haploid induction efficiency (Peloquin et al., 1996; Hermsen and Verdenius, 1973). From these results, it was concluded that the 2x sperm fertilizes the central cell, leaving no sperm to fertilize the egg, which then develops parthenogenetically. However, HI-specific DNA markers in dihaploids or near-dihaploid aneuploids were reported in some studies (Clulow et al., 1991; Waugh et al., 1992; Clulow et al., 1993; Wilkinson et al., 1995; Clulow and Rousselle-Bourgeois, 1997; Pham et al., 2019; Bartkiewicz et al., 2018; Ercolano et al., 2004; Straadt and Rasmussen, 2003; Allainguillaume et al., 1997), but not others (Samitsu and Hosaka, 2002; Amundson et al., 2020). This suggests retention of HI DNA either as chromosomes or segments thereof in an otherwise haploid plant, a diagnostic feature of uniparental chromosome elimination in plants (Zhao et al., 2013; Tan et al., 2015; Kuppu et al., 2015; Gernand et al., 2005; Riera-Lizarazu et al., 1996; Laurie and Bennett, 1986; Kynast et al., 2001; Li et al., 2017; Ishii et al., 2010; Maheshwari et al., 2015). In *Arabidopsis* haploid induction systems, selective instability of the HI chromosomes are observed among the hybrid byproducts (Kuppu et al., 2015; Tan et al., 2015; Maheshwari et al., 2015).

Given the role of dihaploids in efforts to convert potatoes to a diploid inbred crop, the incidental transfer of HI DNA remains a concern, as it has been reported to influence the phenotype of dihaploids (Allainguillaume et al., 1997). The relatively small sample sizes of previous studies, including our previous evaluation of 167 dihaploids that did not identify instances of HI DNA transfer (Amundson et al., 2020), warrant a robust characterization of the frequency and molecular state of incidental HI DNA transfer in primary dihaploids. Furthermore, to our knowledge, no study has investigated genomic stability of triploid and tetraploid hybrids

obtained from potato haploid induction crosses. Toward these ends, we asked two questions: 1) How often, if ever, do potato HIs transmit chromosomes or chromosome fragments to dihaploids? 2) Do dihaploids or hybrid byproducts of potato haploid induction exhibit evidence of genome instability? In this study, we used genome resequencing to search for HI DNA and determine its molecular state in 919 dihaploids and 134 hybrids obtained from potato haploid induction crosses. We found that 8.27% of primary dihaploids were aneuploid, and in about 90% of cases was due to additional missing chromosomes from the non-inducer parent. Eight primary dihaploids exhibited 1-3 additional chromosomes from the HI parent, some of which appeared fragmented due to genome instability. Chromosome breakage in dihaploids and hybrids suggest an association between haploidization and genome instability. However, this instability appears ploidy-dependent: HI chromosomes were fragmented much less often in triploid hybrids than in either dihaploids or tetraploid hybrids. Comparison with tetraploid self-pollinated progeny suggested that HI genome instability observed in tetraploid hybrids was not attributable to pollen, sperm or embryo ploidy *per se*. In summary, our results indicate low levels of HI DNA contribution to dihaploids, and suggest a role for ploidy of the HI gamete in potato haploid induction.

Results

Widespread aneuploidy among primary dihaploids

Potato haploid induction crosses can yield dihaploids, triploid hybrids and tetraploid hybrids, with possible aneuploidy from either parent at any ploidy level (Fig. 3.1). We pollinated 19 tetraploid clones with haploid inducers IVP35, IVP101 or PL4, recorded presence or absence of the inducer-specific embryo spot, and evaluated the ploidy of each plantlet by guard cell

chloroplast counting or flow cytometry (Supplemental Data Set S1.1). Next, 919 putative dihaploids and 134 hybrids were selected for chromosome dosage analysis by low-coverage whole genome sequencing as previously described (Supplemental Data Set S1.2) (Amundson et al., 2020). For each dihaploid, read depth per chromosome was standardized to that dihaploid's tetraploid parent such that values near 1, 2 or 3 corresponded to monosomy, disomy or trisomy, respectively. Aneuploids were then identified as individuals with one or more outlier chromosomes. The analysis was carried out for all 919 dihaploids; representative results are shown for all 229 dihaploids extracted from clone WA.077 (Fig. 3.2A). In this cohort, we identified twenty-seven aneuploids: 25 had one additional chromosome ($2n=2x+1=25$), one had two additional chromosomes ($2n=2x+2=26$) and one had three chromosomes ($2n=2x+3=27$).

Among all sequenced dihaploids, 8.3% were aneuploid. Primary trisomics (i.e., single chromosome aneuploids) composed 91% of the aneuploid class, with the remaining aneuploids consisting of monosomics ($2n=2x-1=23$) and primary trisomics for multiple chromosomes ($2n=2x+2=26$ or $2n=2x+3=27$) (Fig. 3.2B). Each of the 12 homologous chromosomes was recovered as a trisomic. Differences among chromosomes were not significant for either aneuploidy of any type (gains and losses pooled, $p=0.07296$, $df=11$, X^2 test) or chromosome gains only ($p=0.06726$, $df=11$, X^2 test) (Supplemental Fig. S3.1). It is worth noting that flow cytometric analyses did not readily detect aneuploidy in primary dihaploids.

To evaluate the effect of parental genotype on aneuploidy frequency, we grouped dihaploids based on the genotypes of the parents. When grouped by maternal genotype, aneuploidy frequency ranged from 6.4% to 11.3% and differences between maternal genotypes were not significant ($p=0.5987$; $df=6$) (Supplemental Fig. S3.2). When grouped by paternal genotype,

aneuploidy frequency ranged from 7.6% to 10.3% and differences between inducer genotypes were not significant either ($p=0.3626$; $df=2$) (Supplemental Fig. S3.3). Taken together, our data show that approximately 8% of presumed dihaploids were aneuploid, without detectable aneuploidy bias for parental genotype in this material.

Retention of haploid inducer chromosomes

To determine the parental origin of the additional chromosomes in the aneuploid dihaploid progeny, we identified homozygous SNPs between each pair of tetraploid seed parent and haploid inducer. We used these SNPs to calculate the percentage of reads that originated from the haploid inducer across the genome of every dihaploid. If the additional chromosome originated from the haploid inducer, this percentage was expected to be approximately 33%, while it was expected to be close to 0% if all copies originated from the tetraploid parent. Representative SNP dosage plots are shown in Fig. 3.2C. In this population, two of the aneuploids identified in Fig. 3.2A carried chromosomes from the haploid inducer parent. One of these two individuals also exhibiting haploid inducer alleles above background levels on chromosome 1 (Fig. 3.2C). A third individual was not aneuploid according to dosage analysis but showed haploid inducer alleles above background levels on chromosomes 1 and 8 (Fig. 3.2C).

Among all dihaploids, we found 66 aneuploids with additional chromosomes exclusively from the tetraploid parent and 8 with additional chromosomes from the HI (Fig. 3.2D). Two aneuploid lines, MM247 and MM890, were segmental aneuploids with additional HI chromosome segments; the six others showed HI-derived aneuploidy of entire chromosomes (Fig. 3.3; Supplemental Fig. S3.4). We refer to these eight lines as HI addition dihaploids hereafter. All of

the haploid inducer genotypes contributed genetic material to at least one dihaploid, and the frequencies at which they did so were not significantly different ($p=0.7747$; Fisher Exact test). The frequency of aneuploidy frequency was consistent with previous analyses of primary dihaploid populations (Amundson et al., 2020; Pham et al., 2019; Samitsu and Hosaka, 2002; Wagenvoort and Lange, 1975; Hermsen et al., 1970; Hermsen, 1969; Frandsen, 1967), and the low frequency (8/919; 0.87%) of aneuploidy due to additional HI chromosomes agrees with previous results in which chromosomes from the inducer parent were not detected in cohorts of less than 200 individuals (Amundson et al., 2020; Samitsu and Hosaka, 2002; Pham et al., 2019).

Detection of haploid inducer-derived DNA segments in dihaploids

Appearance of inducer DNA fragments shorter than entire chromosomes have also been reported among potato dihaploids (Wilkinson et al., 1995; Pham et al., 2019). Our low coverage sequencing cannot detect segments of this size, but they may be detected with higher coverage. To test whether this type of transfer occurred in our material, we sequenced three HI addition lines to 27-30x coverage and searched for secondary introgressions, i.e., segments of chromosomes other than the trisomic chromosome showing HI-specific SNP alleles. Overall, few markers (0.39-0.69%) were consistent with HI introgression, and HI alleles were underrepresented in allele-specific read depth (Supplemental Fig. S3.5). When considering only putative introgressions covering ≥ 3 adjacent markers, we found that each of the three lines exhibited putative introgressions, with 3-13 events per line and a total of 21 events (Table 3.1). Of these putative introgressions, eight showed identical coordinates in MM247 and MM1114, both of which carried part or all of chromosome 8 from IVP35. One of these putative introgressions, a 1.8 Mb region of chromosome 1, showed linkage to chromosome 8 in seven of our uniparental dihaploid populations (Supplemental Fig. S3.6) and at least three independent

mapping populations (Endelman and Jansky, 2016; Bourke et al., 2015), suggesting they are located on the trisomic-HI derived chromosome, and only appear as introgressions due to errors in the genome assembly used for analysis (DM1-3 version 4.04). Consistent with this prediction, when we aligned each putative introgression to the updated long-read genome assembly (v6.1, (Pham et al., 2020)), we found that eight had an unambiguous top hit to the chromosome corresponding to the HI-derived addition chromosome (Table 3.1). For the remaining thirteen, we used short read alignments to the v4.04 assemblies to locate possible breakpoint junctions. Short read alignments of DM1-3 to itself indicated that these regions were matched uniformly by short reads, but that the putative boundaries were not matched by reads from IVP35, WA.077 and the dihaploid (Table 3.2, Supplemental Data Set S3.1). In conclusion, our analysis did not provide evidence of true introgressions of short HI DNA segments, indicating instead that they can be attributed to structural variation pre-existing haploid induction.

Selective instability of the haploid inducer genome in dihaploids and tetraploid hybrids

Next, we asked whether the hybrid byproducts of potato haploid induction showed signs of genome instability. Seeds with the dominant, inducer-specific embryo spot marker were germinated and analyzed by flow cytometry, yielding 30 triploids and 104 tetraploids. As a control, we included 14 progeny that did not show the nodal banding phenotype, were tetraploid by flow cytometry, and lacked HI alleles in the low coverage sequencing; these are likely self-pollinated progeny of the tetraploid clones. To distinguish novel dosage variants attributable to genome instability from recurring variants likely due to pre-existing structural variation in the parents, each offspring was evaluated in the context of its siblings of the same ploidy (Fig. 3.4A-

B). Aneuploids made up a greater proportion of tetraploid hybrids (>70% vs 22% of triploid hybrids), with the frequency of maternally and paternally derived aneuploidy both increasing (Fig. 3.4C). The per-chromosome rate of HI-derived segmental aneuploidy was significantly lower in the triploids hybrids than the corresponding rate in either dihaploids or tetraploid hybrids, suggesting a greater degree of HI genome instability in dihaploids and tetraploid hybrids (Table 3.2). Chromosome breakage was not observed in tetraploid selfs, suggesting that the instability of HI-derived chromosomes seen in tetraploid hybrids was not a consequence of 2x pollen and/or sperm *per se* (Table 3.2). Relative to triploid hybrids, tetraploid hybrids showed a strong and highly significant increase in the incidence of HI-derived genome instability ($p < 0.001$; log odds 95% CI 2.62247-1.501144) (Fig. 3.4C), a difference driven by more frequent segmental aneuploidy of HI-derived chromosomes (Fig. 3.4D). Most tetraploid hybrids exhibited no more than two novel CNV of each parental genome, indicating that genome instability is not restricted to few exceptional tetraploid hybrids, but is pervasive (Fig. 3.4E). In conclusion, our data suggest that HI-derived chromosomes are selectively unstable in dihaploids and tetraploids, suggesting a specific role of 2x HI sperm in potato haploid induction.

Tetraploid hybrids produced by first meiotic division restitution of the haploid inducer

The rate of restitution sperm, but not of 2n pollen, appears to be associated with potato haploid induction (Montelongo-Escobedo and Rowe, 1969; Peloquin et al., 1996; Dongyu et al., 1995). We next asked how each 2x sperm was formed, and whether any one mechanism was more likely to be associated with HI genome instability. From the high-coverage sequencing of each dihaploid addition line, we derived the HI haplotypes of the additional chromosome(s) in each

line (Fig. 3.5A), which we refer to as H' hereafter. Using these haplotypes, we then used the low-pass sequencing of 134 hybrids to genotype the centromeres of the chromosomes contributed by the HI. As a control, we analyzed triploid hybrids and found the expected transmission of a single HI haplotype through the centromere and into the chromosome arms (Supplemental Fig. S3.7), indicating that our centromeric HI haplotype phasing was robust. For tetraploid hybrids, the HI-contributed sequences at centromere-linked markers were expected to be heterozygous (show ~25% H' allele) if derived from 2n first division restitution (FDR) pollen, but homozygous (show either ~0% or ~50% HI allele) if derived from either 2n second division restitution (SDR) pollen or 2x restitution sperm (Fig. 3.5B). Among 78 tetraploid hybrids, all but five showed HI heterozygosity at the centromeres, implicating FDR as the dominant mechanism of hybrid formation (Fig. 3.5C; Supplemental Fig. S3.8). Among the five hybrids with the SDR or RS pattern, one showed ~50% H' allele of Cen11, but this individual was disomic for chromosome 11, with both maternal homologs missing (Supplemental Fig. S3.9); together, these results are also consistent with FDR. Of the remaining four hybrids, all of which were derived from IVP101, two showed signs of HI genome instability and two did not. In conclusion, FDR hybrids predominated among tetraploids, with minor contributions possibly from SDR gametes or restitution sperm. On the other hand, no mechanism was uniquely associated with HI genome instability.

Discussion

Multiple studies investigating the presence of haploid inducer genetic material in potato dihaploids have come to different conclusions (Amundson et al., 2020; Pham et al., 2019; Bartkiewicz et al., 2018; Ercolano et al., 2004; Straadt and Rasmussen, 2003; Samitsu and

Hosaka, 2002; Clulow and Rousselle-Bourgeois, 1997; Allainguillaume et al., 1997; Wilkinson et al., 1995; Clulow et al., 1993; Waugh et al., 1992; Clulow et al., 1991). Documenting the existence and extent to which this type of DNA transfer occurs is critical for basic understanding of haploid induction, as well as the use of primary dihaploids for diploid potato breeding. Here, we sequenced a cohort of 919 dihaploids and found that 0.87% of primary dihaploids contained 1-3 chromosomes or chromosome fragments from the HI. Introgressions of smaller HI DNA segments were also detected, but in about a third of cases, could be explained by errors in reference genome assembly, while the remaining putative introgressions could not be robustly confirmed. Haploid inducer chromosomes were generally stable in triploid hybrids, but unstable in tetraploid hybrids, most of which were products of first division restitution pollen.

We documented the occasional appearance of small HI segments (0.5 to few kb). These could represent small translocations or gene conversion derived from the HI genome before elimination. The evidence in support of their presence is robust because it is based on a continuous haplotype that encompasses multiple SNPs. However, in 8 out of 21 instances these segments were physically linked to another HI chromosome present in these samples. For example, identical HI segments were present in HI addition lines MM247 and MM1114, which were both trisomic for chromosome 8. Part of chromosome 8 was previously identified as translocated or misassembled in at least two genetic mapping populations (Endelman and Jansky, 2016; Bourke et al., 2015) as well as in our dihaploid populations. Therefore, these short introgressions are physically part of the trisomic chromosome and map to a different location of the reference assembly, possibly due to genome assembly artifacts or pre-existing translocations in the HI relative to the reference genome. The remaining 13 were flanked by poorly mapping

sequences and could not be anchored to a chromosome. While introgression could not be ruled out in these cases, the probability that these are segmental recombination events is low.

The presence of HI DNA in some dihaploids could be explained if, after formation of a hybrid genome, the HI genome was imperfectly eliminated. Acceptance of this hypothesis would imply that uniparental genome elimination is the mechanism that generates haploids and that the dihaploids with no HI DNA contamination underwent perfect elimination of the HI genome.

There are, however, at least two alternative explanations. First, that all or a fraction of the perfect dihaploids result from parthenogenesis. This would require two independent mechanisms of haploid induction to work in the potato system, and seems implausible. Second, that the mechanism of haploid induction entails incomplete gamete fusion. A defective sperm may thus deliver both egg-activating factors and, occasionally, chromosomes, but fail to carry out proper karyogamy. This mechanism has not been explicitly proposed, to our knowledge, but may explain observations from the maize haploid induction system, where in addition to phospholipase, mutation of fusogenic proteins both enhance HI rate, and induce haploids in the wild-type phospholipase background (Jacquier et al., 2020; Zhong et al., 2020, 2019).

What fates are possible for the HI genome? The genome delivered by sperms formed by the tetraploid selfs were stable. We could assess HI genome integrity in HI-contaminated dihaploids, and in triploid and tetraploid hybrids. Dihaploids that had inherited and maintained HI chromosomes shared high instability of the HI genome with tetraploid hybrids. Triploids, on the other hand, did not. This demonstrates that the HI genome can be inherited and maintained with fidelity, and that instability is not intrinsic to the formation of hybrid zygotes. The tetraploids result from hybridization of a $1n(=2x)$ egg sac with $2n(=2x)$ sperm. We infer that either during

formation of the 2n sperm or upon fertilization, the HI genome becomes unstable. The instability displayed by tetraploid hybrids could be related to that displayed by HI-containing dihaploids. This provides a potential explanation for the long-standing proposal that 2n sperm triggers haploid induction (Montelongo-Escobedo and Rowe, 1969; Montezuma-de-Carvalho, 1967; Wangenheim et al., 1960). These studies documented the formation of 2x sperm from restitution of the generative cell mitosis in pollen of HIs and suggested a connection to haploid induction. Genome maintenance may become compromised during formation or growth of 2n pollen resulting in a fragmented genome that is incompetent for replication and subject to elimination.

It is also possible that instability is unrelated to genome elimination. We explored the nature of 2n sperms in the HI crosses by analyzing the HI contribution in tetraploid hybrids. Restitution of second mitosis predicts homozygosity of 2n sperm. Instead we found heterozygosity indicative of meiotic First Division Restitution (FDR) in most cases. This agrees with a previous study of microsporogenesis in IVP35 that reported moderate frequencies of parallel spindles (16.88-26.13%) and fused spindles (1.29-17.65%) that results in ~29% dyads (Ramanna, 1979). Both parallel and fused spindles are associated with FDR (d'Erfurth et al., 2009; De Storme and Geelen, 2011; Li et al., 2010; Peloquin et al., 1999). However, this leaves unanswered the role of 2x sperm in HI. It demonstrates, however, that the instability observed in the tetraploid hybrids may be connected to FDR. This instability could result from missegregation during meiosis (Umbreit et al., 2020; Ly et al., 2019, 2017), but its relation to HI is mysterious.

The frequency of aneuploids among primary dihaploids (8.27%) was within the range of previously studied dihaploid populations (1.5-11.4%) (Amundson et al., 2020; Pham et al., 2019;

Samitsu and Hosaka, 2002; Wagenvoort and Lange, 1975; Hermsen et al., 1970; Hermsen, 1969; Frandsen, 1967). Based on the availability of robust polymorphic markers, we identified HI DNA in 8 individuals out of 919, 6 with whole chromosomes and 2 with large fragments. That aneuploids with HI chromosomes make up less than 1% of primary dihaploids is good news for diploid potato breeding, as our results suggest that most primary dihaploids will be free of residual HI DNA. More often, primary dihaploids carry additional maternal chromosomes, possibly due to meiotic nondisjunction in the autotetraploid, female parent. Little evidence is available on the cytology of female meiosis in potato and other autoployploids (Ramsey and Schemske, 2002), but in male meiosis, high frequencies of univalents and multivalents at metaphase I (He et al., 2018; Swaminathan, 1954b) and unbalanced chromosome sets at metaphase II (Swaminathan, 1954a) have been reported. Regardless of origin, the immediate and potentially lasting impacts of aneuploidy (Henry et al., 2010) are likely adverse and best left avoided. Unlike the primary trisomics of *Datura*, *Arabidopsis* and tomato, which show diagnostic phenotypes (Blakeslee, 1922; Koornneef and Van der Veen, 1983; Steinitz-Sears, 1963; Rick and Barton, 1954), primary trisomics could not be readily detected by phenotype alone in primary dihaploids of potato (Wagenvoort and Lange, 1975; Hermsen et al., 1970), indicating that cytological or genetic assays would be required to detect it.

In conclusion, using a large-scale approach, we examined the genomes of 1,053 progeny from HI crosses determining that, regardless of parental genotype, a small but definite fraction of dihaploids display paternal HI contribution. This large-scale study provides the solid evidence needed to interpret previous studies, and calibrates the expectations for potato HI crosses. In addition, we made an unexpected observation. The HI genome, which is stable when inherited by

triploid hybrids, displays selective instability both in tetraploid hybrids and in dihaploids. The interpretation and meaning of these findings are still open. At a minimum, they indicate that genome stability is compromised in the HI 2n pollen. They also reinforce the hypothesis that 2n pollen may be required for HI, suggesting future lines of investigation to elucidate mechanisms contributing to this unusual, but highly relevant phenomenon.

Materials and Methods

Plant material

Primary dihaploids and hybrids were obtained from 19 tetraploid clones (Supplemental Table S3.2) *via* pollination with haploid inducers IVP101 (Hutten et al., 1993), IVP35 (Hermsen and Verdenius, 1973) or PL4 (also known as CIP596131.4; Ordoñez et al, in prep) in greenhouses located at the CIP's experimental station in the Peruvian Andes (3,216 masl, -12.01039, -75.22411). Flower buds of the pistillate parents were emasculated and pollinated with HI pollen the following day. All haploid inducers are homozygous for a dominant embryo spot that facilitates the detection of hybrids (Hermsen and Verdenius, 1973; Hutten et al., 1993). Seeds were extracted from mature fruit, recorded for presence or absence of the embryo spot, and germinated on soil. The ploidy of each established seedling was determined by either chloroplast counting as described in (Amundson et al., 2020) or flow cytometric measurement of nuclear DNA content against maternal and paternal parents as standards. Refer to Supplemental Data Set S3.2 for an expanded description of plant material.

Flow cytometry

Approximately 50-60 mg of greenhouse-grown leaf tissue was harvested from each sample and homogenized in 500 μ l of LB01 buffer (Doležel et al., 1989) and left to rest for 1 minute. 250 μ l of homogenate was passed through a 20 μ m filter (Partec 04-0042-2315) into tubes containing 12 μ l of 1mg/ml propidium iodide and 2.5 μ l of 5mg/ml RNase. Samples were incubated in the dark for 5 minutes and analyzed in an Accuri C6 flow cytometer (BD biosciences) with the following filter configurations: a) FL-1 530/14-nm bandpass filter, b) FL-2 585-20nm bandpass filter and c) FL-3 670-nm longpass filter. Threshold levels were set to 10,000 for forward scatter (FSC) with a secondary threshold of 1,000 for FL-2 (Galbraith et al., 2011).

Whole genome resequencing

Genomic DNA was extracted from young leaflets as previously described (Ghislain M., Zhang D. P., Herrera, M. R., 1999). For each sample, approximately 750ng of genomic DNA was sheared to an average size of 300bp as previously described (Amundson et al., 2020). Libraries were constructed using all sheared input DNA with KAPA Hyper Prep kit (cat. No KK8504) with half-scale reactions used throughout the protocol, custom 8bp index adapters, and amplification cycles as described in Supplemental Dataset S3.3. Libraries were sequenced on Illumina HiSeq 4000 or NovaSeq 6000 platforms at the University of California, Davis DNA Technologies Core, Vincent Coates Genome Sequencing Laboratory, or University of California San Francisco Center for Advanced Technologies, as specified in Supplemental Data Set S3.3. Libraries were demultiplexed using custom Python scripts available on our laboratory website ([allprep-12.py](http://comailab.genomecenter.ucdavis.edu/index.php/Barcoded_data_preparation_tools); http://comailab.genomecenter.ucdavis.edu/index.php/Barcoded_data_preparation_tools).

Publicly available sequencing reads from (Pham et al., 2017), (Hardigan et al., 2017) and (Amundson et al., 2020) were retrieved from NCBI Sequence Read Archive and incorporated in subsequent analyses.

Variant calling

Adapter and low quality sequences were trimmed from raw reads using Cutadapt v1.15 (Martin, 2011), retaining reads ≥ 40 nt in length. Trimmed reads were aligned to the DM1-3 v4.04 reference assembly, including DM1-3 chloroplast and mitochondrion sequences, using BWA mem (v0.7.12r1039) and default settings (Li, 2013). Alignments were further processed to remove PCR duplicates, soft clip one mate of overlapping read pairs, remove read pairs with mates aligning to different chromosomes, and locally realign indels, as previously described (Amundson et al., 2020). Processed alignments were then used as input for joint variant calling and genotyping with FreeBayes (version 1.3.2) (Garrison and Marth, 2012) with minimum mapping quality 20, base quality 20, Hardy-Weinberg priors off, and up to 4 alleles considered per variant, and all other parameters left at the default setting.

Initially, we genotyped a subset of parental clones with deep whole genome sequencing available from this study or from previous studies (Hardigan et al., 2017; Pham et al., 2017), which we designated “Cohort A” in Supplemental Data Set S3.2. Raw variants were filtered as follows: NUMALT == 1, CIGAR == 1X, QUAL ≥ 20 , MQM ≥ 50 , MQMR ≥ 50 , |MQM - MQMR| < 10, RPPR ≤ 20 , RPP ≤ 20 , EPP ≤ 20 , EPPR ≤ 20 , SAP ≤ 20 , SRP ≤ 20 . For each pair of tetraploid parent and haploid inducer represented in the offspring of Cohort A, we identified loci with read depth within 1.5 times the genome-wide median of each parent, at least 10 supporting reads, and homozygous genotype calls for different alleles in the two parents. Each list of parental SNPs

was used to provisionally determine chromosome dosage (see ‘Chromosome Dosage Analysis’ below) for all offspring of Cohort A. To determine parental origin of aneuploidy in offspring for which sequencing reads from the tetraploid parent were not available, we tested the possibility of using pooled reads from multiple dihaploids produced from the same parent instead. Specifically, we tested the effect of substituting pooled low-coverage sequencing from dihaploids for that same tetraploid parent at the SNP calling step, and tested if we could recapitulate the observations obtained using SNPs taken directly from the tetraploid parent. As a proof of concept, we pooled low-coverage alignments from 205 dihaploids of WA.077 at the variant calling step and repeated all downstream analysis of Cohort A samples. Upon obtaining acceptable results, pooled alignments from low coverage dihaploids extracted from C93.154 (n=237), 93.003 (n=73), C91.640 (n=79), LR00.014 (n=110), LR00.022 (n=51), LR00.026 (n=51), WA.077 (n=205), and all deeply sequenced HI addition lines were included along with Cohort A samples for the variant calling and genotyping reported in the manuscript.

Chromosome dosage analysis

Read alignments from low coverage dihaploids and hybrids were filtered for mapping quality ≥ 10 and counted in non-overlapping 1Mb bins using bedtools (version 2.27.1). We calculated the fraction of all aligned reads that mapped to a chromosome, normalized this fraction to the corresponding fraction of a family-specific control sample (controls specified in Supplemental Data Set S3.2) and scaled the standardized coverage values to the expected ploidy state based on flow cytometry results. Putative aneuploids were identified as outliers with a standardized coverage value of ≥ 3 standard deviations from the within-family all-chromosome mean. In some families, segregation of pre-existing deletions on chromosome 12 resulted in a high rate of false positive trisomy and monosomy calls. False positives of this nature are listed in Supplemental

Data Set S3.2. Individuals exhibiting a false positive signal on chromosome 12 were not recorded as aneuploid, unless they also exhibited aneuploidy of another chromosome type. To infer parental origin of numerical and structural aneuploidies, parental SNPs were identified for each combination of tetraploid parent and haploid inducer as described above. For each low-coverage dihaploid or hybrid, allele-specific read depth was then tallied at homozygous parent-informative SNP loci in non-overlapping 4Mb bins, and bins with fewer than 30 reads covering all informative loci within a bin were withheld from analysis.

High resolution analysis of parental DNA contribution

For each HI addition line, genotype data were recorded as T=tetraploid parent H=haploid inducer. For high stringency filtering, loci were removed from consideration if any of the following criteria were met: i) one or more reads matched the HI allele in the tetraploid parent, ii) three or more reads matched the HI allele in the dihaploid pool, iii) excessive read depth (greater than the mean depth plus four standard deviations greater than the mean depth) was observed in either parent or the dihaploid at hand (Li, 2014), iv) HI allele depth was < 6 in the dihaploid at hand or v) the HI allele represented $< 15\%$ of the total reads at a locus in the dihaploid at hand. For regions of interest, we viewed short read alignments to the reference genome DM1-3 using the Integrative Genome Viewer (Thorvaldsdottir et al., 2013).

Low-pass haplotype analysis

Biallelic SNP loci with homozygous genotype calls for either allele in WA.077, heterozygous genotype calls in IVP35 and heterozygous with a single dose of the HI-specific allele (i.e., 0/0/1 if the called tetraploid genotype was 0/0/0/0 and 0/1/1 if the called genotype was 1/1/1/1) were used to define phased alleles of H'. For each tetraploid hybrid, we then calculated the depth of

reads with H' and non-H' alleles at all retained loci, aggregated counts across the DM1-3 coordinates defined as recombination-suppressed centromeres by Bourke et al. (2015) or by non-overlapping 4Mb bins, and reported the ratio of reads matching H' reads to H' + non-H' reads.

Statistical analyses

Proportion of aneuploids among dihaploids by female parent

Euploidy and aneuploidy were treated as discrete outcomes, and counts of each category were evaluated for statistical significance using the `prop.test()` function in R version 3.6.2 (R Core Team 2019). Only families with 30 or more dihaploids were included in the analysis.

Fisher exact counts of HI dihaploid introgression events by haploid inducer

Appearance of HI-derived chromosomes in an otherwise dihaploid plant was treated as a binary outcome and used to construct a 2x3 contingency table with each HI genotype. Only dihaploids for which we had sufficient SNP information to determine chromosome parental origin were considered. This included dihaploids from the following tetraploid parents: 93.003 (CIP390637.1), Atlantic (CIP800827), C01.020 (CIP301023.15), C91.640 (CIP3888615.22), C93.154 (CIP392820.1), Desiree (CIP800048), LR00.014 (CIP300056.33), LR00.022 (CIP300072.1), LR00.026 (CIP300093.14) and WA.077 (CIP397077.16). This table was used to conduct a Fisher Exact test in R using the function `fisher.test()`.

Linkage disequilibrium from dosage variable states

For each dihaploid, standardized coverage values and bin dosage states were derived for non-overlapping 1Mb bins of the reference genome as previously described (Amundson et al., 2020).

Fisher Exact tests were then carried out between pairs of dosage states to assess linkage

disequilibrium between bins. For example, assume that both Bin1 and Bin100 have three dosage states: standardized coverages 1, 2 and 3. To test whether Bin1-State1 was correlated with Bin100-CN3, the following four dihaploid sets were compared in a 2x2 contingency table: *observed in Bin1-State1 : observed not in Bin1-State1, expected in Bin1-State1 : expected not in Bin1-State1*, where the expectation was derived from the assumption of complete independence. Self-comparison and reciprocal comparisons were removed, and the remaining comparisons were controlled at false discovery rate (FDR) = 0.05 unless otherwise noted. Chromosomal bins in statistically significant linkage disequilibrium (LD) with one another for any state at that pair of bins were displayed in an LD matrix.

Logistic regression model for incidence of paternal (maternal) genome instability in triploid and tetraploid hybrids

For each hybrid, we determined the incidence and parental origin of whole-chromosome and segmental aneuploidy tetraploid hybrids as previously described for potato dihaploids (Amundson et al., 2020). Instability of the paternal genome was treated as a binary outcome and used in a logistic regression model with ploidy, genotype of the HI parent and aneuploidy of maternal chromosomes included as predictor variables. Contribution of the maternal genome to aneuploidy of HI-derived chromosomes was determined from the maternal aneuploidy term in the model. Pairwise differences between ploidy levels were evaluated with Tukey multiple test correction. To evaluate the stability of maternal chromosomes, maternal aneuploidy was used as the binary response variable and paternal aneuploidy was incorporated as a predictor variable. Effects of paternal aneuploidy on maternal aneuploidy, as well as ploidy-dependent effects were evaluated as described above.

Data Availability

All sequencing data generated in this study is currently being deposited at NCBI Sequence Read Archive under a Project ID PRJNA699631. IVP101 whole genome sequencing was retrieved from NCBI Sequence Read Archive project ID PRJNA408137. Whole genome sequencing of cv. “Atlantic” was retrieved from NCBI Sequence Read Archive project ID PRJNA287438. Code for read preprocessing, variant calling, and chromosome dosage analysis is available on https://github.com/kramundson/MM_manuscript. Supplemental Data Sets S3.1, S3.2 and S3.3 are available on Dryad at doi:10.25338/B8JS8D.

Acknowledgements

We thank Hannele Lindqvist-Kreuzer for guidance and support of the CIP-based research and the CIP's greenhouse technicians who performed most of the crossings and plant maintenance. This work was supported by the National Science Foundation Plant Genome Integrative Organismal Systems (IOS) Grant 1444612 (Rapid and Targeted Introgression of Traits via Genome Elimination) to L.C.

Author Contributions

L.C. and E.-H.T. conceived experiments, K.R.A., B.O., E.H.T., M.B., A.K., I.M.H. and L.C. designed experiments, K.R.A., B.O., M.S. and L.C. performed experiments, K.R.A., B.O., M.S., and L.C. performed data analysis, K.R.A., M.L.N., E.H.T., I.M.H. and L.C. interpreted data, B.O., M.S., M.B., and A.K. contributed reagents and materials, and K.R.A. and L.C. wrote the paper with input from all authors.

Conflict of Interest Statement

The authors declare no conflict of interest in this study.

Literature Cited

- Allainguillaume, J., Wilkinson, M.J., Clulow, S.A., and Barr, S.N.R.** (1997). Evidence that genes from the male parent may influence the morphology of potato dihaploids. *Theor. Appl. Genet.* **94**: 241–248.
- Amundson, K.R., Ordoñez, B., Santayana, M., Tan, E.H., Henry, I.M., Mihovilovich, E., Bonierbale, M., and Comai, L.** (2020). Genomic Outcomes of Haploid Induction Crosses in Potato (*Solanum tuberosum* L.). *Genetics* **214**: 369–380.
- Bartkiewicz, A.M., Chilla, F., Terefe-Ayana, D., Lübeck, J., Strahwald, J., Tacke, E., Hofferbert, H.-R., Linde, M., and Debener, T.** (2018). Maximization of Markers Linked in Coupling for Tetraploid Potatoes via Monoparental Haploids. *Front. Plant Sci.* **9**: 620.
- Blakeslee, A.F.** (1922). Variations in *Datura* Due to Changes in Chromosome Number. *Am. Nat.* **56**: 16–31.
- Bourke, P.M., Voorrips, R.E., Visser, R.G.F., and Maliepaard, C.** (2015). The Double-Reduction Landscape in Tetraploid Potato as Revealed by a High-Density Linkage Map. *Genetics* **201**: 853–863.
- Clulow, S.A. and Rousselle-Bourgeois, F.** (1997). Widespread introgression of *Solanum phureja* DNA in potato (*S. tuberosum*) dihaploids. *Plant Breed.* **116**: 347–351.
- Clulow, S.A., Wilkinson, M.J., and Burch, L.R.** (1993). *Solanum phureja* genes are expressed in the leaves and tubers of aneusomatic potato dihaploids. *Euphytica* **69**: 1–6.
- Clulow, S.A., Wilkinson, M.J., Waugh, R., Baird, E., Demaine, M.J., and Powell, W.** (1991). Cytological and molecular observations on *Solanum phureja*-induced dihaploid potatoes. *Theor. Appl. Genet.* **82**: 545–551.
- De Storme, N. and Geelen, D.** (2011). The *Arabidopsis* mutant *jason* produces unreduced first division restitution male gametes through a parallel/fused spindle mechanism in meiosis II. *Plant Physiol.* **155**: 1403–1415.
- Doležel, J., Binarová, P., and Lcretti, S.** (1989). Analysis of nuclear DNA content in plant cells by flow cytometry. *Biol. Plant.* **31**: 113–120.
- Dongyu, Q., Dewei, Z., Ramanna, M.S., and Jacobsen, E.** (1995). A comparison of progeny from diallel crosses of diploid potato with regard to the frequencies of 2n-pollen grains. *Euphytica* **92**: 313–320.
- Endelman, J.B. and Jansky, S.H.** (2016). Genetic mapping with an inbred line-derived F2 population in potato. *Theor. Appl. Genet.* **129**: 935–943.

- Ercolano, M.R., Carputo, D., Li, J., Monti, L., Barone, A., and Frusciante, L.** (2004). Assessment of genetic variability of haploids extracted from tetraploid ($2n = 4x = 48$) *Solanum tuberosum*. *Genome* **47**: 633–638.
- d’Erfurth, I., Jolivet, S., Froger, N., Catrice, O., Novatchkova, M., and Mercier, R.** (2009). Turning meiosis into mitosis. *PLoS Biol.* **7**: e1000124.
- Frandsen, N.O.** (1967). Haploid production in a potato breeding material with intensive back crossing to wild species. *Zuchter* **37**: 120–134.
- Gabert, A.C.** (1963). Factors influencing the frequency of haploids in the common potato (*Solanum tuberosum* L.).
- Galbraith, D.W., Janda, J., and Lambert, G.M.** (2011). Multiparametric analysis, sorting, and transcriptional profiling of plant protoplasts and nuclei according to cell type. *Methods Mol. Biol.* **699**: 407–429.
- Garrison, E. and Marth, G.** (2012). Haplotype-based variant detection from short-read sequencing. arXiv [q-bio.GN].
- Gernand, D., Rutten, T., Varshney, A., Rubtsova, M., Prodanovic, S., Brüß, C., Kumlehn, J., Matzk, F., and Houben, A.** (2005). Uniparental Chromosome Elimination at Mitosis and Interphase in Wheat and Pearl Millet Crosses Involves Micronucleus Formation, Progressive Heterochromatinization, and DNA Fragmentation. *Plant Cell* **17**: 2431–2438.
- Ghislain M., Zhang D. P., Herrera, M. R. ed** (1999). *Molecular Biology Laboratory Protocols Plant Genotyping: Training Manual* (International Potato Center: Lima, Peru).
- Hanneman, R.E. and Peloquin, S.J.** (1968). Ploidy levels of progeny from diploid-tetraploid crosses in the potato. *Am. Potato J.* **45**: 255–261.
- Hanneman, R.E. and Ruhde, R.W.** (1978). Haploid extraction in *Solanum tuberosum* group Andigena. *Am. J. Potato Res.* **55**: 259–263.
- Hardigan, M.A., Laimbeer, F.P.E., Newton, L., Crisovan, E., Hamilton, J.P., Vaillancourt, B., Wiegert-Rininger, K., Wood, J.C., Douches, D.S., Farré, E.M., Veilleux, R.E., and Buell, C.R.** (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. U. S. A.* **114**: E9999–E10008.
- He, L., Braz, G.T., Torres, G.A., and Jiang, J.** (2018). Chromosome painting in meiosis reveals pairing of specific chromosomes in polyploid *Solanum* species. *Chromosoma* **127**: 505–513.
- Henry, I.M., Dilkes, B.P., Miller, E.S., Burkart-Waco, D., and Comai, L.** (2010). Phenotypic consequences of aneuploidy in *Arabidopsis thaliana*. *Genetics* **186**: 1231–1245.
- Hermesen, J.G.T.** (1969). Induction of haploids and aneuploids in colchicine-induced tetraploid *Solanum chacoense* Bitt. *Euphytica* **18**: 183–189.
- Hermesen, J.G.T. and Verdenius, J.** (1973). Selection from *Solanum tuberosum* group phureja of genotypes combining high-frequency haploid induction with homozygosity for embryo-spot. *Euphytica* **22**: 244–259.
- Hermesen, J.G.T., Wagenvoort, M., and Ramanna, M.S.** (1970). ANEUPLOIDS FROM NATURAL

- AND COLCHICINE-INDUCED AUTOTETRAPLOIDS OF SOLANUM. *Can. J. Genet. Cytol.* **12**: 601–613.
- Hutten, R.C.B., Scholberg, E.J.M.M., Huigen, D.J., Hermsen, J.G.T., and Jacobsen, E.** (1993). Analysis of dihaploid induction and production ability and seed parent x pollinator interaction in potato. *Euphytica* **72**: 61–64.
- Ishii, T., Ueda, T., Tanaka, H., and Tsujimoto, H.** (2010). Chromosome elimination by wide hybridization between Triticeae or oat plant and pearl millet: pearl millet chromosome dynamics in hybrid embryo cells. *Chromosome Res.* **18**: 821–831.
- Jackson, M.T., Rowe, P.R., and Hawkes, J.G.** (1978). Crossability relationships of andean potato varieties of three ploidy levels. *Euphytica* **27**: 541–551.
- Jacquier, N.M.A., Gilles, L.M., Pyott, D.E., Martinant, J.-P., Rogowsky, P.M., and Widiez, T.** (2020). Puzzling out plant reproduction by haploid induction for innovations in plant breeding. *Nat Plants* **6**: 610–619.
- Jansky, S.H. et al.** (2016). Reinventing Potato as a Diploid Inbred Line–Based Crop. *Crop Sci.* **56**: 1412–1422.
- Koorneef, M. and Van der Veen, J.H.** (1983). Trisomics in *Arabidopsis thaliana* and the location of linkage groups. *Genetica* **61**: 41–46.
- Kuppu, S., Tan, E.H., Nguyen, H., Rodgers, A., Comai, L., Chan, S.W.L., and Britt, A.B.** (2015). Point Mutations in Centromeric Histone Induce Post-zygotic Incompatibility and Uniparental Inheritance. *PLoS Genet.* **11**: e1005494.
- Kynast, R.G. et al.** (2001). A Complete Set of Maize Individual Chromosome Additions to the Oat Genome. *Plant Physiology* **125**: 1216–1227.
- Laurie, D.A. and Bennett, M.D.** (1986). Wheat × maize hybridization. *Can. J. Genet. Cytol.* **28**: 313–316.
- Li, H.** (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio.GN].
- Li, H.** (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**: 2843–2851.
- Lindhout, P., Meijer, D., Schotte, T., Hutten, R.C.B., Visser, R.G.F., and van Eck, H.J.** (2011). Towards F1 Hybrid Seed Potato Breeding. *Potato Res.* **54**: 301–312.
- Li, X., Meng, D., Chen, S., Luo, H., Zhang, Q., Jin, W., and Yan, J.** (2017). Single nucleus sequencing reveals spermatid chromosome fragmentation as a possible cause of maize haploid induction. *Nat. Commun.* **8**: 991.
- Li, Y., Shen, Y., Cai, C., Zhong, C., Zhu, L., Yuan, M., and Ren, H.** (2010). The type II *Arabidopsis* formin14 interacts with microtubules and microfilaments to regulate cell division. *Plant Cell* **22**: 2710–2726.
- Ly, P., Brunner, S.F., Shoshani, O., Kim, D.H., Lan, W., Pyntikova, T., Flanagan, A.M., Behjati, S., Page, D.C., Campbell, P.J., and Cleveland, D.W.** (2019). Chromosome segregation errors generate

- a diverse spectrum of simple and complex genomic rearrangements. *Nat. Genet.* **51**: 705–715.
- Ly, P., Teitz, L.S., Kim, D.H., Shoshani, O., Skaletsky, H., Fachinetti, D., Page, D.C., and Cleveland, D.W.** (2017). Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by non-homologous end joining. *Nat. Cell Biol.* **19**: 68–75.
- Maheshwari, S., Tan, E.H., West, A., Franklin, F.C.H., Comai, L., and Chan, S.W.L.** (2015). Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genet.* **11**: e1004970.
- Marks, G.E.** (1966). The enigma of triploid potatoes. *Euphytica* **15**: 285–290.
- Martin, M.** (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12.
- Montelongo-Escobedo, H. and Rowe, P.R.** (1969). Haploid induction in potato: Cytological basis for the pollinator effect. *Euphytica* **18**: 116–123.
- Montezuma-de-Carvalho, J.** (1967). The effect of N₂O on pollen tube mitosis in styles and its potential significance for inducing haploidy in potato. *Euphytica* **16**: 190–198.
- Peloquin, S.J., Boiteux, L.S., and Carputo, D.** (1999). Meiotic mutants in potato. Valuable variants. *Genetics* **153**: 1493–1499.
- Peloquin, S.J., Gabert, A.C., and Ortiz, R.** (1996). Nature of “Pollinator” Effect in Potato (*Solanum tuberosum* L.) Haploid Production. *Ann. Bot.* **77**: 539–542.
- Pham, G.M., Braz, G.T., Conway, M., Crisovan, E., Hamilton, J.P., Laimbeer, F.P.E., Manrique-Carpintero, N., Newton, L., Douches, D.S., Jiang, J., Veilleux, R.E., and Buell, C.R.** (2019). Genome-wide Inference of Somatic Translocation Events During Potato Dihaploid Production. *Plant Genome* **12**.
- Pham, G.M., Hamilton, J.P., Wood, J.C., Burke, J.T., Zhao, H., Vaillancourt, B., Ou, S., Jiang, J., and Buell, C.R.** (2020). Construction of a chromosome-scale long-read reference genome assembly for potato. *Gigascience* **9**.
- Pham, G.M., Newton, L., Wiegert-Rininger, K., Vaillancourt, B., Douches, D.S., and Buell, C.R.** (2017). Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. *Plant J.* **92**: 624–637.
- Ramanna, M.S.** (1979). A re-examination of the mechanisms of 2n gamete formation in potato and its implications for breeding. *Euphytica* **28**: 537–561.
- Ramsey, J. and Schemske, D.W.** (2002). Neopolyploidy in Flowering Plants. *Annu. Rev. Ecol. Syst.* **33**: 589–639.
- Rick, C.M. and Barton, D.W.** (1954). Cytological and Genetical Identification of the Primary Trisomics of the Tomato. *Genetics* **39**: 640–666.
- Riera-Lizarazu, O., Rines, H.W., and Phillips, R.L.** (1996). Cytological and molecular characterization of oat x maize partial hybrids. *Theor. Appl. Genet.* **93**: 123–135.
- Samitsu, Y. and Hosaka, K.** (2002). Molecular marker analysis of 24- and 25-chromosome plants

obtained from *Solanum tuberosum* L. subsp. *andigena* ($2n = 4x = 48$) pollinated with a *Solanum phureja* haploid inducer. *Genome* **45**: 577–583.

Spooner, D.M., Ghislain, M., Simon, R., Jansky, S.H., and Gavrilenko, T. (2014). Systematics, Diversity, Genetics, and Evolution of Wild and Cultivated Potatoes. *Bot. Rev.* **80**: 283–383.

Steinitz-Sears, L.M. (1963). Chromosome Studies in *Arabidopsis Thaliana*. *Genetics* **48**: 483–490.

Straadt, I.K. and Rasmussen, O.S. (2003). AFLP analysis of *Solanum phureja* DNA introgressed into potato dihaploids. *Plant Breed.* **122**: 352–356.

Swaminathan, M.S. (1954a). MICROSPOROGENESIS IN SOME COMMERCIAL POTATO VARIETIES. *J. Hered.* **45**: 265–272.

Swaminathan, M.S. (1954b). Nature of Polyploidy in Some 48-Chromosome Species of the Genus *Solanum*, Section *Tuberarium*. *Genetics* **39**: 59–76.

Tan, E.H., Henry, I.M., Ravi, M., Bradnam, K.R., Mandakova, T., Marimuthu, M.P., Korf, I., Lysak, M.A., Comai, L., and Chan, S.W. (2015). Catastrophic chromosomal restructuring during genome elimination in plants. *Elife* **4**.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**: 178–192.

Umbreit, N.T., Zhang, C.-Z., Lynch, L.D., Blaine, L.J., Cheng, A.M., Tourdot, R., Sun, L., Almubarak, H.F., Judge, K., Mitchell, T.J., Spektor, A., and Pellman, D. (2020). Mechanisms generating cancer genome complexity from a single cell division error. *Science* **368**.

Wagenvoort, M. and Lange, W. (1975). The production of aneuidihaploids in *Solanum tuberosum* L. group *Tuberosum* (the common potato). *Euphytica* **24**: 731–741.

Wangenheim, K.H., Peloquin, S.J., and Hougas, R.W. (1960). Embryological investigations on the formation of haploids in the potato (*Solanum tuberosum*). *Molecular and General*.

Waugh, R., Baird, E., and Powell, W. (1992). The use of RAPD markers for the detection of gene introgression in potato. *Plant Cell Rep.* **11**: 466–469.

Wilkinson, M.J., Bennett, S.T., Clulow, S.A., Allainguillaume, J., Harding, K., and Bennett, M.D. (1995). Evidence for somatic translocation during potato dihaploid induction. *Heredity* **74 (Pt 2)**: 146–151.

Zhao, X., Xu, X., Xie, H., Chen, S., and Jin, W. (2013). Fertilization and uniparental chromosome elimination during crosses with maize haploid inducers. *Plant Physiol.* **163**: 721–731.

Zhong, Y. et al. (2020). A DMP-triggered in vivo maternal haploid induction system in the dicotyledonous *Arabidopsis*. *Nat Plants* **6**: 466–472.

Zhong, Y. et al. (2019). Mutation of *ZmDMP* enhances haploid induction in maize. *Nat Plants* **5**: 575–580

Figures

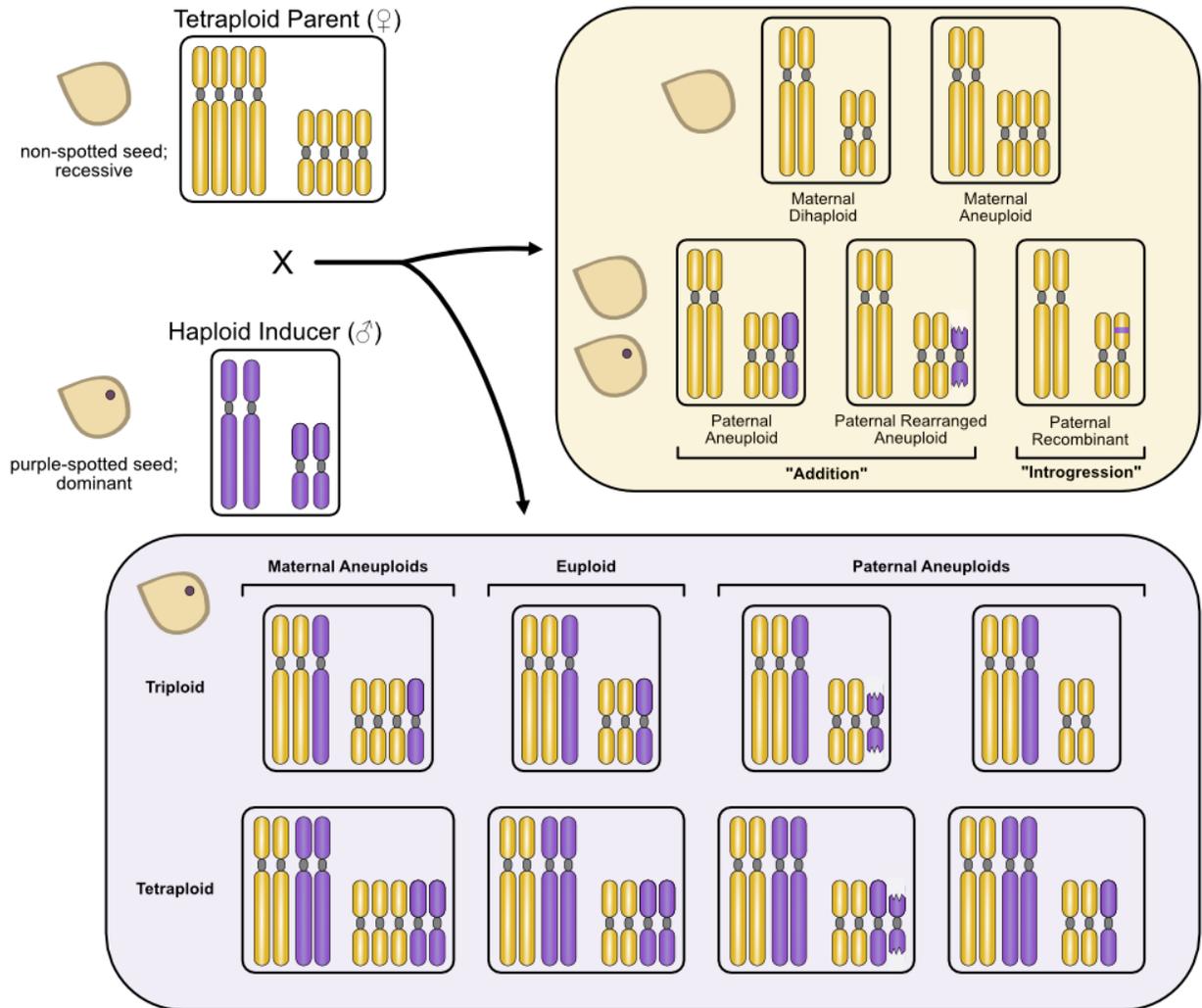


Figure 3.1. Possible outcomes of potato haploid induction crosses. The haploid-inducing pollinator genotypes used in this study were homozygous for the dominant embryo spot trait. Presence or absence of haploid inducer DNA in the ensuing progeny is expected to manifest as presence or absence of the embryo spot. Progeny from spotted seeds include hybrids that may be aneuploid as well. If potato haploid induction is due to post-zygotic elimination of paternal chromosomes, then occasional failure to eliminate all paternal DNA is expected to result in additional paternal chromosomes, intact or rearranged (addition) and/or integration of paternal DNA segments into maternal chromosomes (introgression). Spotted seeds presumably contain the haploid inducer genome and may be triploid or tetraploid depending on the ploidy of the sperm that took part in fertilization. Diploid sperm may result from unreduced pollen or blockage of generative cell division (Montelongo-Escobedo and Rowe, 1969).

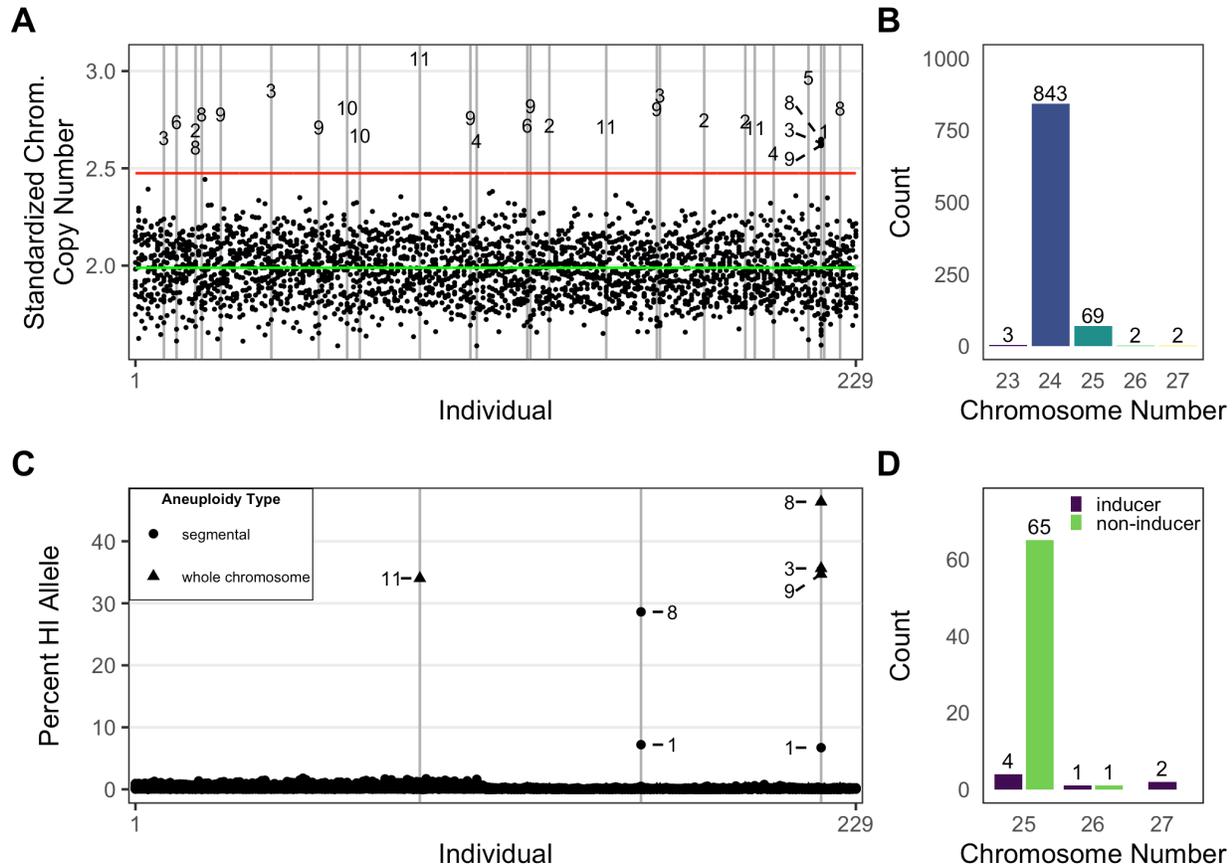


Figure 3.2. Incidence and parental origin of aneuploidy among putative dihaploids. **A)** Standardized chromosome coverage of 229 dihaploids (inferred from flow cytometry) that were extracted from CIP315047 (WA.077). Each individual is displayed along the X-axis, with the stack of 12 points at each coordinate along the X-axis corresponding to the estimated copy number of the 12 potato chromosomes. The green line corresponds to the population all-chromosome mean, and the red line a cutoff of 3 standard deviations greater than the mean, which was our criterion for calling whole-chromosome aneuploidy. Outliers in this distribution correspond to additional chromosomes, all of which are numbered by homolog. **B)** Count of dihaploids by chromosome number inferred from low pass sequencing for all dihaploids evaluated for chromosome dosage (n=1,001) in this study. **C)** Per-chromosome haploid inducer allele contribution of the 229 flow-cytometry confirmed dihaploids shown in panel A. Each individual is displayed along the X axis as a stack of 12 points, with each point corresponding to the haploid inducer allele contribution of one of the 12 chromosomes. Outliers are numbered by chromosome. Chromosomes identified as outliers in panel A are labeled as whole-chromosome aneuploids; those not identified as outliers are labeled as segmental aneuploids. **D)** Parental origin of chromosomal deficiencies and excesses for all near-dihaploid aneuploids analyzed for parental origin in this study (n=73). All compound trisomics resulted in inheritance of multiple additional chromosomes from the same parent, i.e., a 26-chromosome individual exhibited additional chromosomes from either the maternal or paternal parent, but not both.

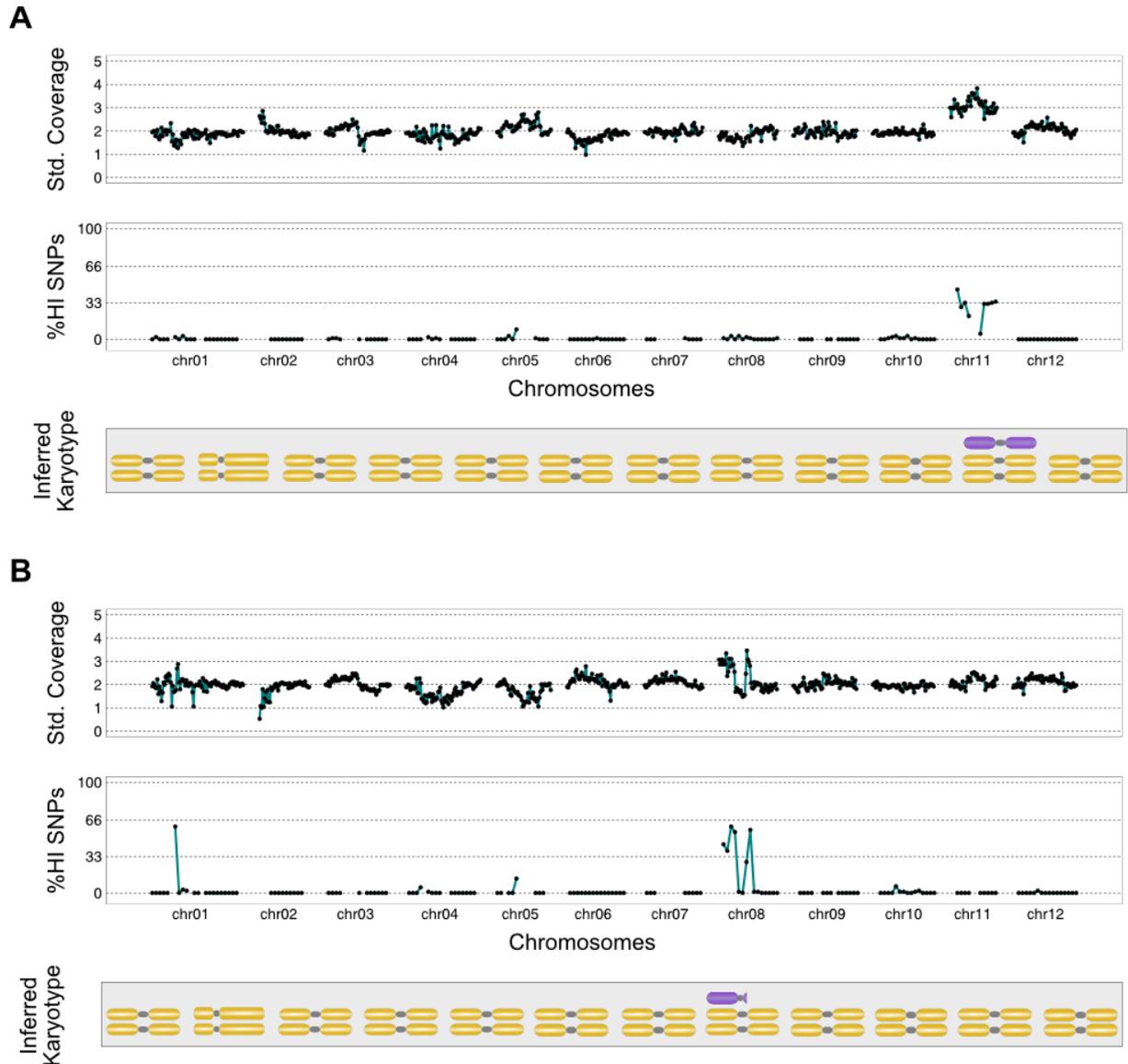
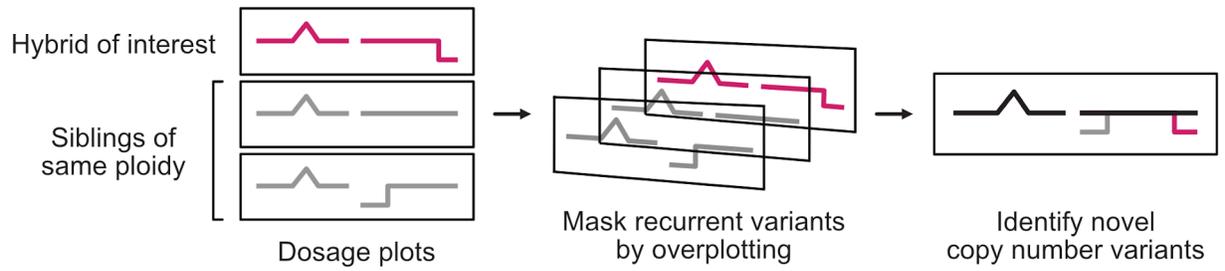
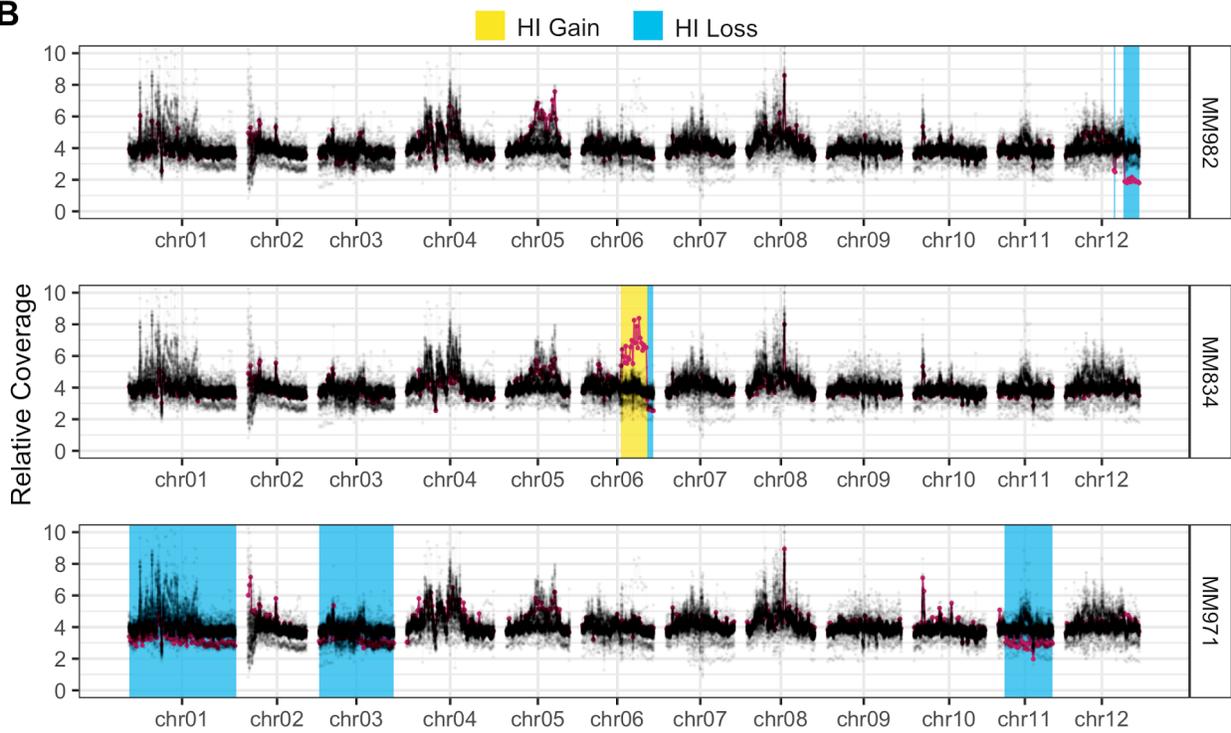
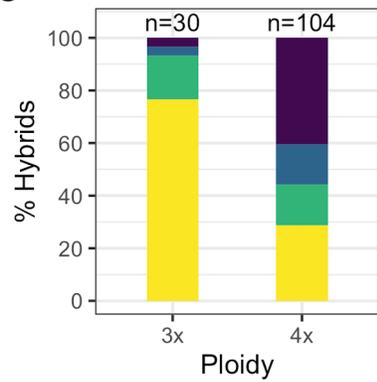
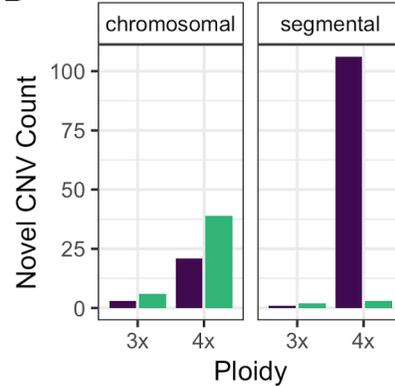
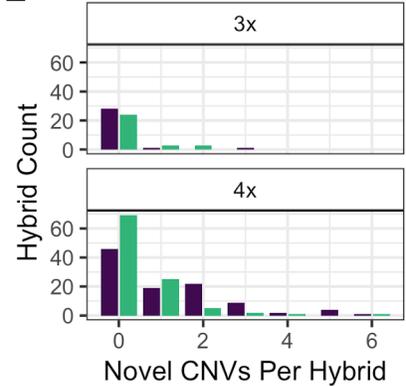


Figure 3.3. Paternal genomic contributions to maternal dihaploids. *In silico* karyotypes of trisomic dihaploids demonstrate the presence of haploid inducer DNA. **A)** Dihaploid MM246, with additional chromosome 11 from haploid inducer IVP35. **B)** Dihaploid MM247, with segmental aneuploidy of chromosome 8 from haploid inducer IVP35. The HI segmental addition on chromosome 1 cosegregates with chromosome 8 in each of seven dihaploid populations we evaluated in this study, suggesting that the two loci are physically linked and that potato is either polymorphic for Chr8-1 translocation, or that this is an assembly error in the reference genome.

A**B****C****D****E**

Aneuploidy Parental Origin ■ None (euploid) ■ NonHI ■ Both ■ HI

Figure 3.4. Haploid inducer (HI) genome instability in tetraploid potato hybrids. Chromosomal variation was investigated by dosage analysis in triploid and tetraploid hybrids. **A)** Schematic of dosage plot generation for a hybrid cohort. An individual displaying novel dosage variation is identified as an outlier track (pink) compared to common structural variation (gray) in overlaying dosage plots obtained by plotting siblings of the same ploidy. Parental origin of each novel variant is inferred from allele-specific read depth at parent-informative SNP loci. **B)** Overlay plots of the same sibling hybrid family with novel copy number variations (CNV) highlighted. Regions corresponding to DNA gains or losses are shaded with yellow or blue backgrounds, respectively. In these examples, all novel CNV are attributable to gained or lost haploid inducer DNA. **C)** Increased paternal aneuploidy in near-tetraploid vs. near-triploid offspring. Combined data from all cohorts in this study. **D)** Paternally derived segmental variation is preponderant in tetraploid hybrids. The bars display counts of aneuploidy according to paternal origin and ploidy of individuals. They also display the counts of whole chromosome aneuploidy vs segmental aneuploidy. As only haploid inducer chromosome breaks are considered in this panel, both classes may exhibit whole chromosome aneuploidy of either parent and segmental aneuploidy of non-inducer chromosomes in addition to HI chromosome breakage. **E)** Number of novel CNV events per hybrid, subdivided by parental origin, showing that haploid inducer genome instability in hybrids is not restricted to few individuals.

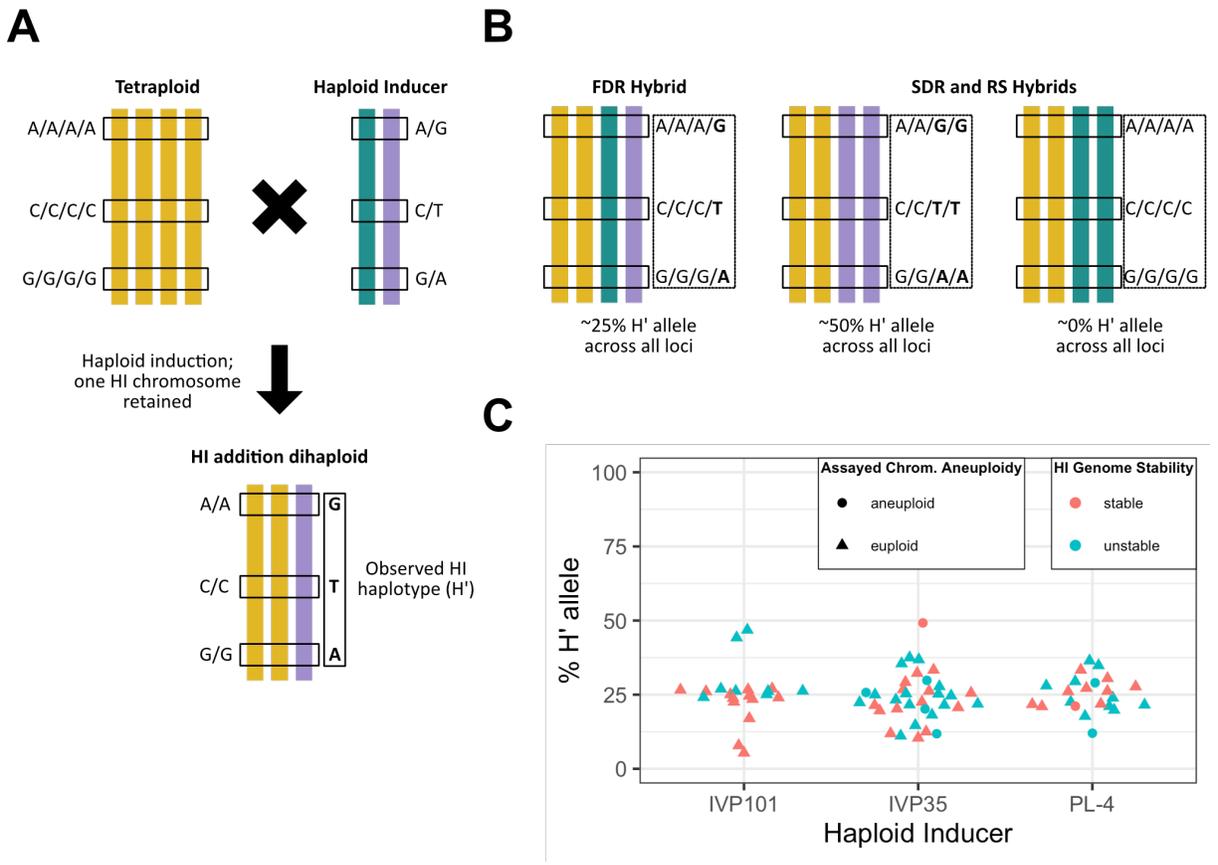


Figure 3.5. Tetraploid hybrid formation in potato haploid induction. **A)** Biallelic SNP loci used for analysis are homozygous in the tetraploid parent and heterozygous but unphased in the HI. The retained HI chromosome(s) in each HI addition dihaploid represents a phased HI-derived haplotype. To avoid confounding effects from crossovers, only expectations and data for the non-recombining centromere is shown. **B)** Expected representation of HI haplotype alleles in a non-recombining region of tetraploid hybrids. For all tetraploid hybrids, read information adjacent loci is binned and the percentage of reads with H' alleles is calculated. If tetraploids are the product of first division restitution (FDR) of the HI, then H' alleles are expected to appear in 25% of all binned reads. If tetraploids are the product of second division restitution (SDR) or restitution sperm (RS), the expected percentage could be 50% or 0% depending on which HI haplotype was inherited. **C)** Percentage of H' allele in the centromeres of tetraploid hybrids. Each point corresponds to the percentage of H' allele among reads spanning the non-recombining region of a chromosome of one tetraploid hybrid. Chromosome 8 was used to assess IVP101 hybrids. Chromosome 10 was used to assess PL4 hybrids. Chromosome 11 was used to assess IVP35 hybrids.

Tables

Table 3.1. Putative introgressions of haploid inducer (HI) DNA segments in dihaploid potatoes. For each HI addition dihaploid, the trisomic chromosome and reference genome coordinates of putative segmental introgressions are shown.

HI Addition Dihaploid	Tetraploid Parent	HI Parent	Trisomic Chromosome(s)	Segments of HI alleles	Seen in multiple dihaploids ^a	Best hit in v6.1 assembly (if different from original chromosome) ^b
MM246	WA.077	IVP35	Chr11	Chr01:45392940-45400039	No	
				Chr05:15510198-15510533	No	
				Chr07:7833610-7844546	No	
				Chr07:7880027-7880413	No	
MM247	WA.077	IVP35	Chr08	Chr01:24228918-25308116	Yes	Chr08
				Chr01:25309455-26003606	Yes	Chr08
				Chr01:38172625-38174004	Yes	Chr08
				Chr07:9467446-9467874	Yes	Chr08
MM1114	WA.077	IVP101	Chr03, Chr08, Chr09	Chr01:24228918-26003606	Yes	Chr08
				Chr01:38172625-38174004	Yes	Chr08
				Chr01:41494766-41499334	No	
				Chr01:67944376-67945914	No	Chr08
				Chr01:84519844-84521775	No	
				Chr05:280661-281197	No	
				Chr05:901127-901352	No	
Chr05:15812350-15814025	No					

				Chr07:5953200-5953784	No	
				Chr07:9467446-9467874	Yes	Chr08
				Chr10:52856295-52859220	No	
				Chr10:54984736-54988488	No	
				Chr12:55689534-55690262	No	

^a Coordinates of putative introgression common to two HI addition dihaploids.

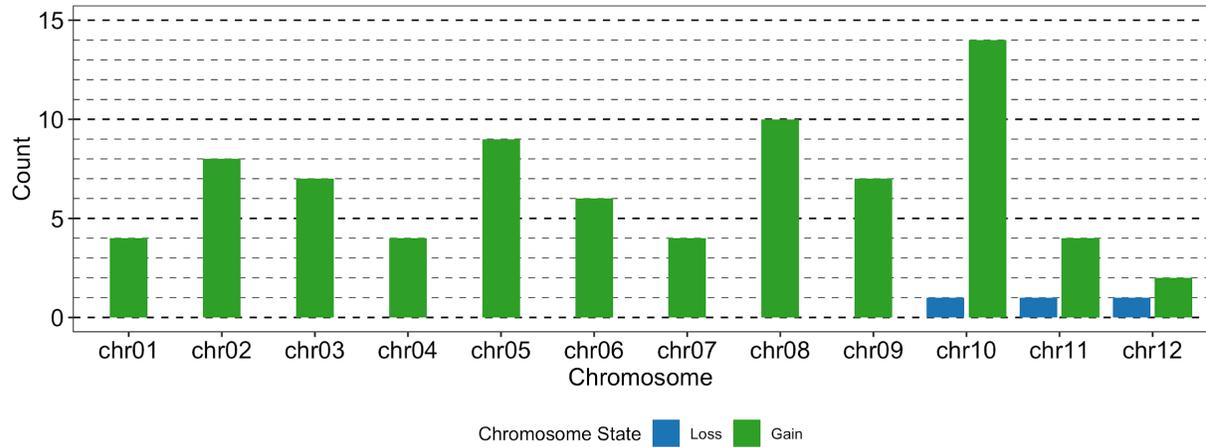
^b Top hit to v6.1 assembly was to a trisomic HI-derived chromosome with full query coverage and >99.5% nucleotide identity.

Table 3.2. Frequency of chromosome breakage among progeny of potato haploid induction crosses.
 For each progeny class, maternal and paternal chromosomes were recorded as appearing in an intact or fragmented state. The number and frequency of chromosomes of each type (intact vs. fragmented) were then grouped by rogeny ploidy and parental origin.

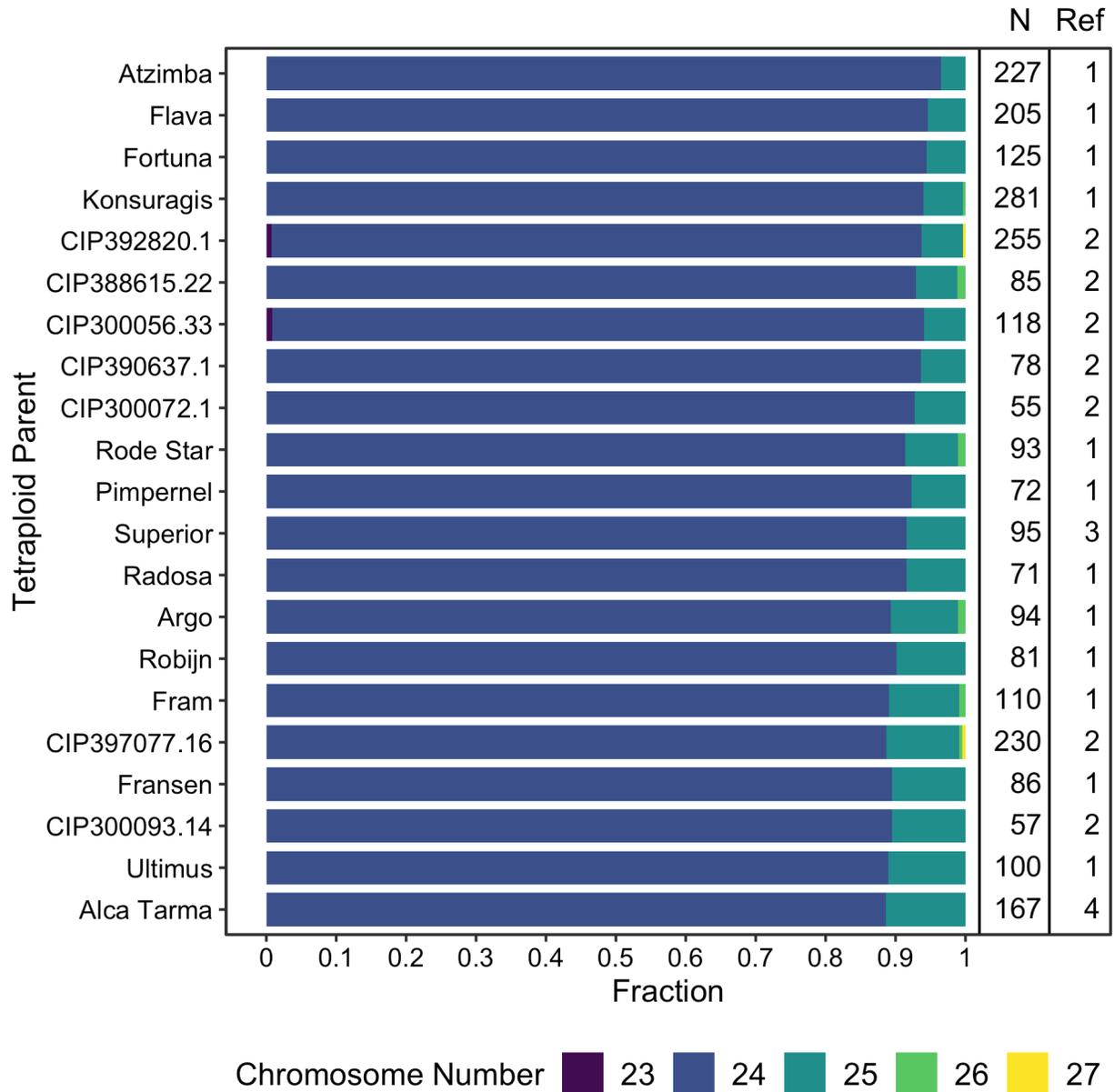
	HI cross progeny	Individuals scored	Chromosomes		<i>P</i> Fisher Exact		
			Intact	Fragmented (%)			
Paternal Chromosomes	Dihaploids	8	11	2 (15.38)	0.0033		
	3x hybrids	30	360	1 (0.28)		<0.00001	0.1117
	4x hybrids	104	2,366	109 (4.40)			
Maternal Chromosomes	Dihaploids	918	22,091	5 (0.02)	0.0186		
	3x hybrids	30	714	2 (0.28)		0.3135	0.0387
	4x hybrids	104	2,470	3 (0.12)			
Putative tetraploid selfs		14	673	0 (0)			

Supplemental Material

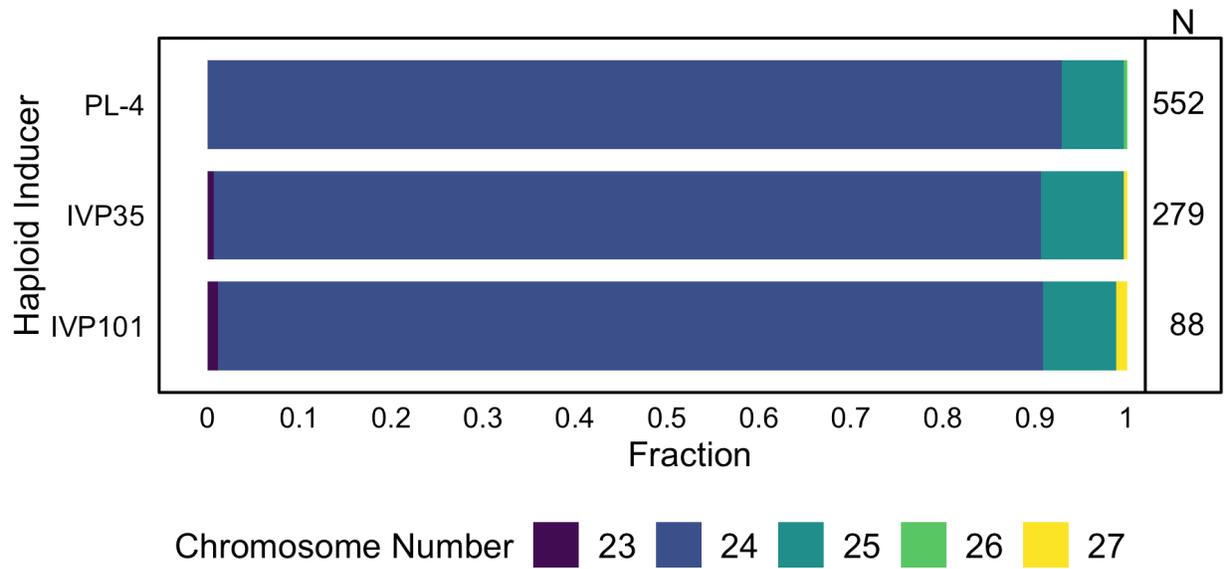
Supplemental Figures



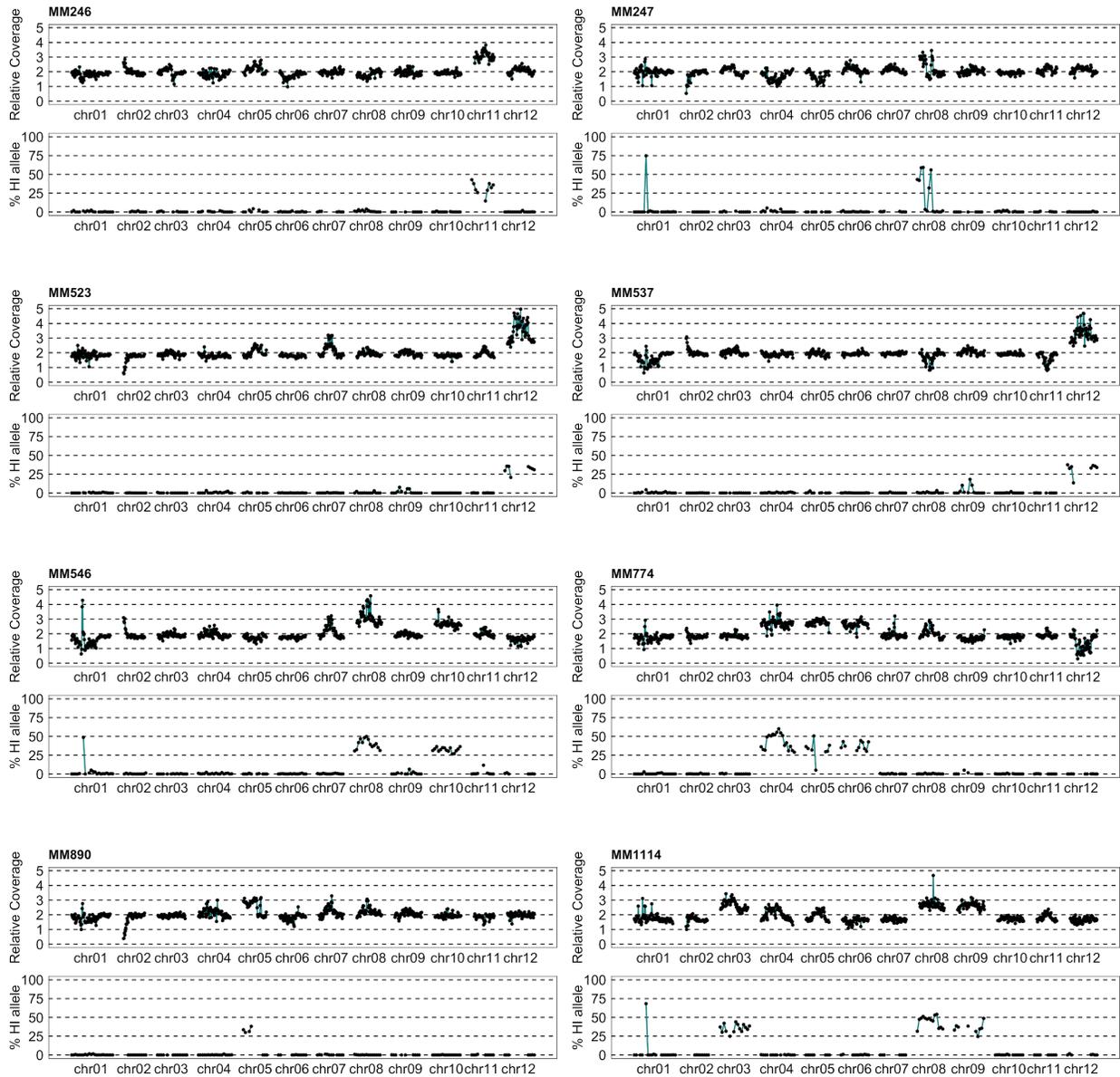
Supplemental Figure S3.1. Histogram of aneuploidy by homologous chromosome affected and state loss or gain among 74 potato dihaploids. Instances of chromosomal gain or loss are counted by homologous chromosome and state of loss or gain.



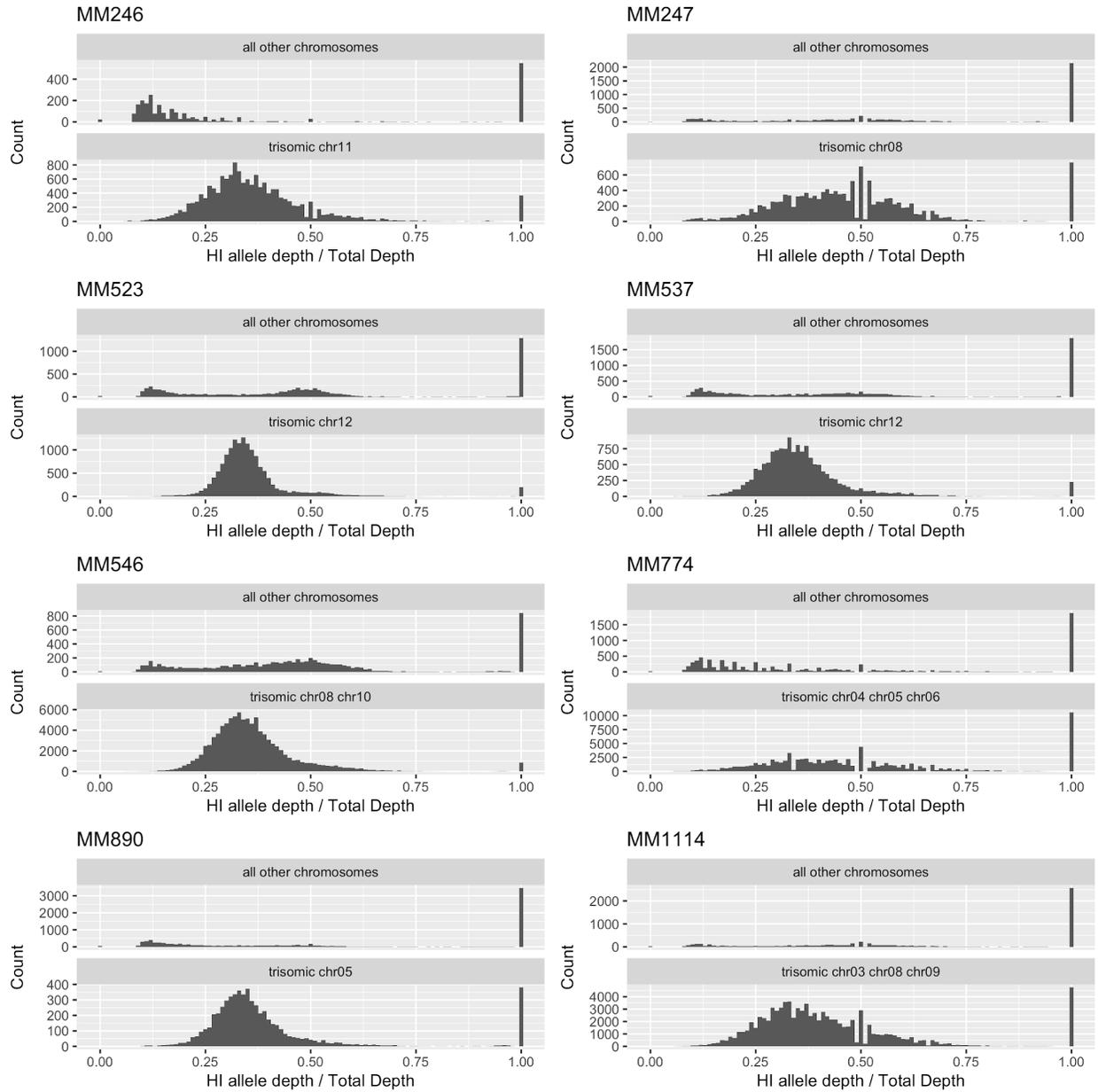
Supplemental Figure S3.2. Aneuploidy frequency among putatively uniparental dihaploid potatoes, partitioned by maternal parent. For each of 21 tetraploid potato clones, the frequencies of euploid and various aneuploid karyotypes among extracted dihaploid are reported by chromosome number. The dihaploid population size (N) and corresponding study are listed for reference (Ref) on the right. References: 1) (Wagenvoort and Lange, 1975), 2) this study, 3) (Pham et al., 2019), 4) (Amundson et al., 2020). Only populations with 50 or more dihaploids are shown.



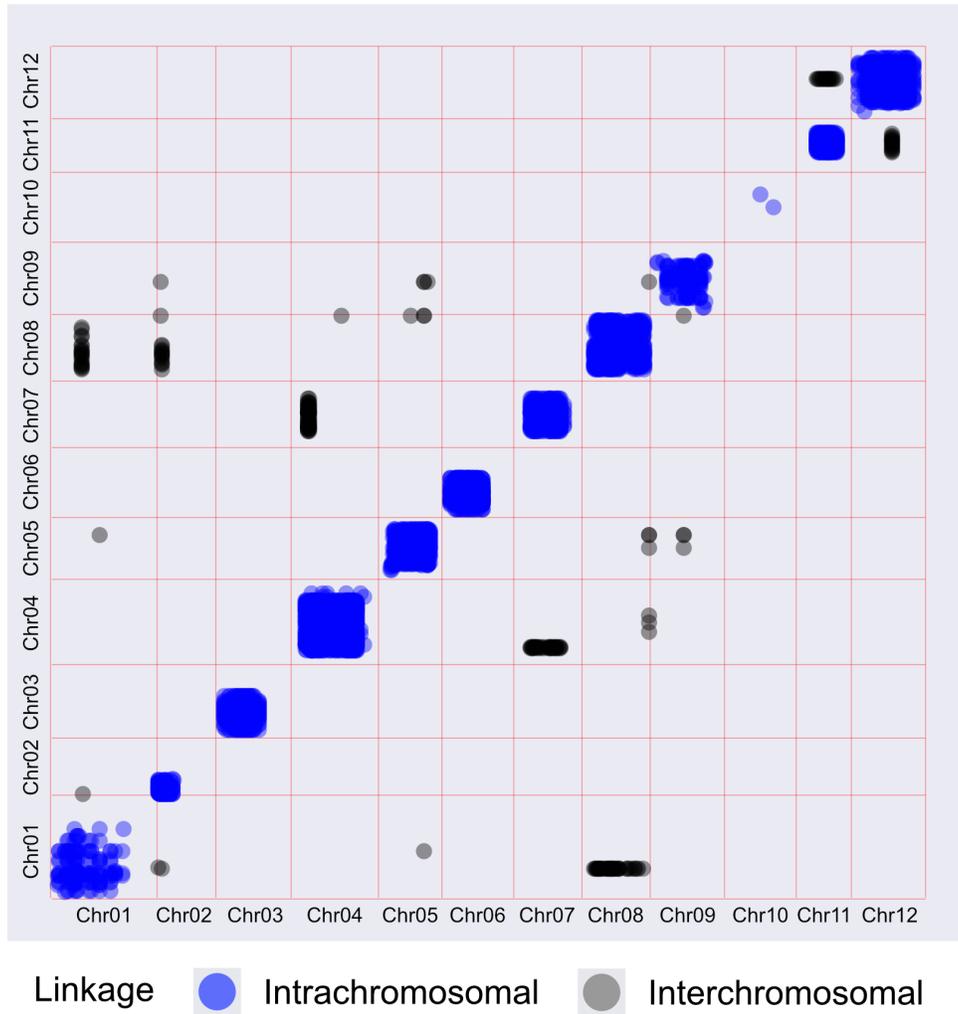
Supplemental Figure S3.3. Aneuploidy frequency among putatively uniparental dihaploid potatoes, partitioned by paternal parent. For each of 3 haploid inducers, the frequencies of euploid and various aneuploid karyotypes among extracted dihaploid are reported by chromosome number. The dihaploid population size (N) is listed on the right.



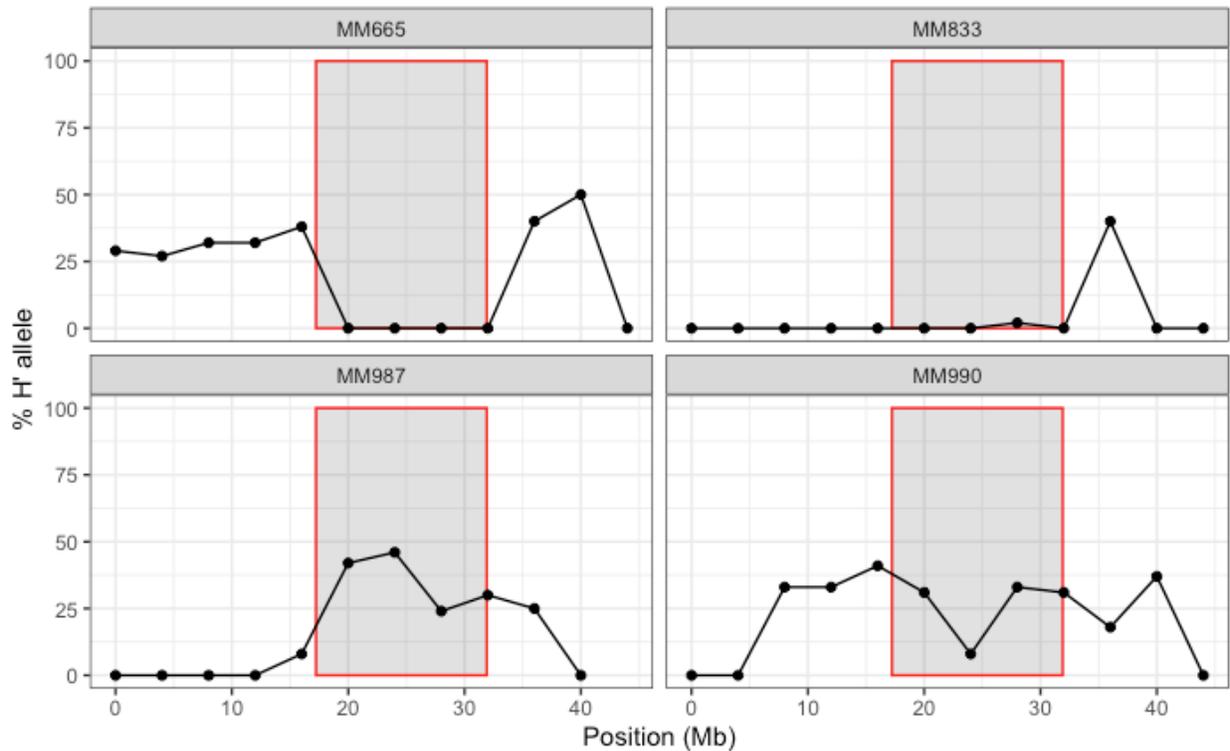
Supplemental Figure S3.4. Chromosome dosage and parental allele dosage plots of eight haploid inducer addition dihaploids. Each dihaploid is represented as a pair of vertically stacked plots. The upper plot displayed relative read coverage to each dihaploid's tetraploid parent. Each point corresponds to the relative coverage of a non-overlapping 1Mb bin of the reference genome. The lower plot displays the percentage of HI-specific allele at all SNP loci identified in non-overlapping 4Mb bins of the reference genome.



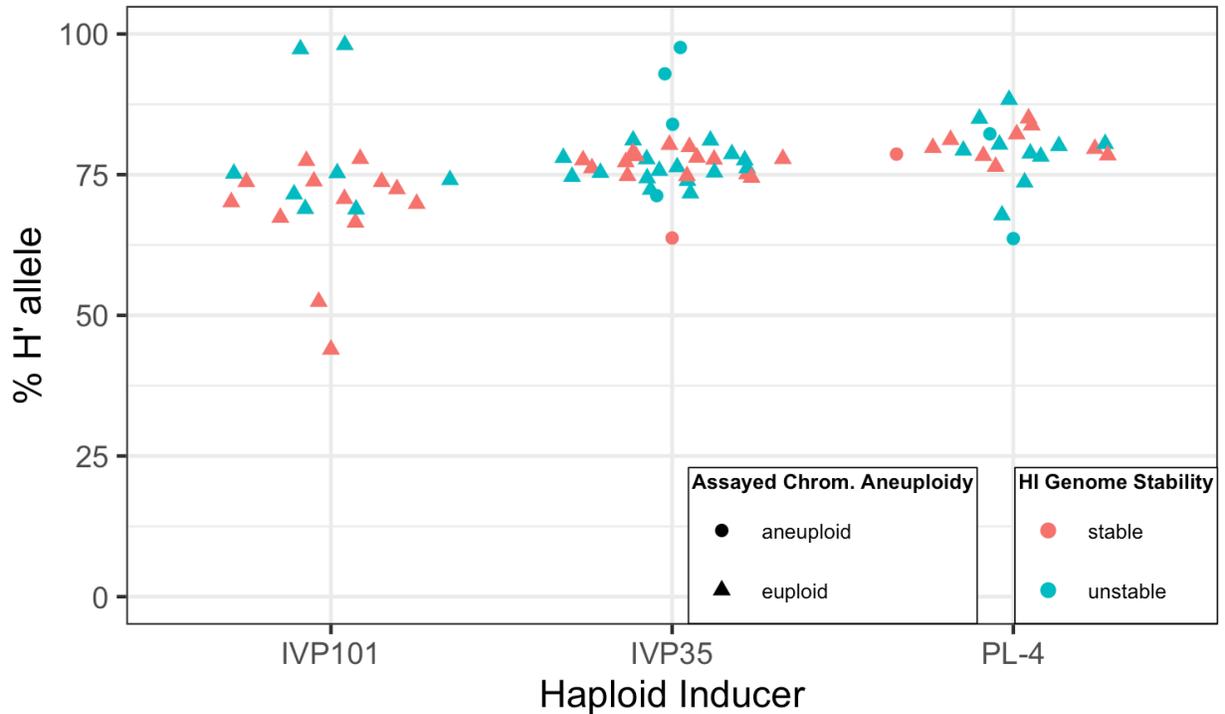
Supplemental Figure S3.5. Histograms of haploid inducer allele representation at putative introgression loci in eight HI addition dihaploids.



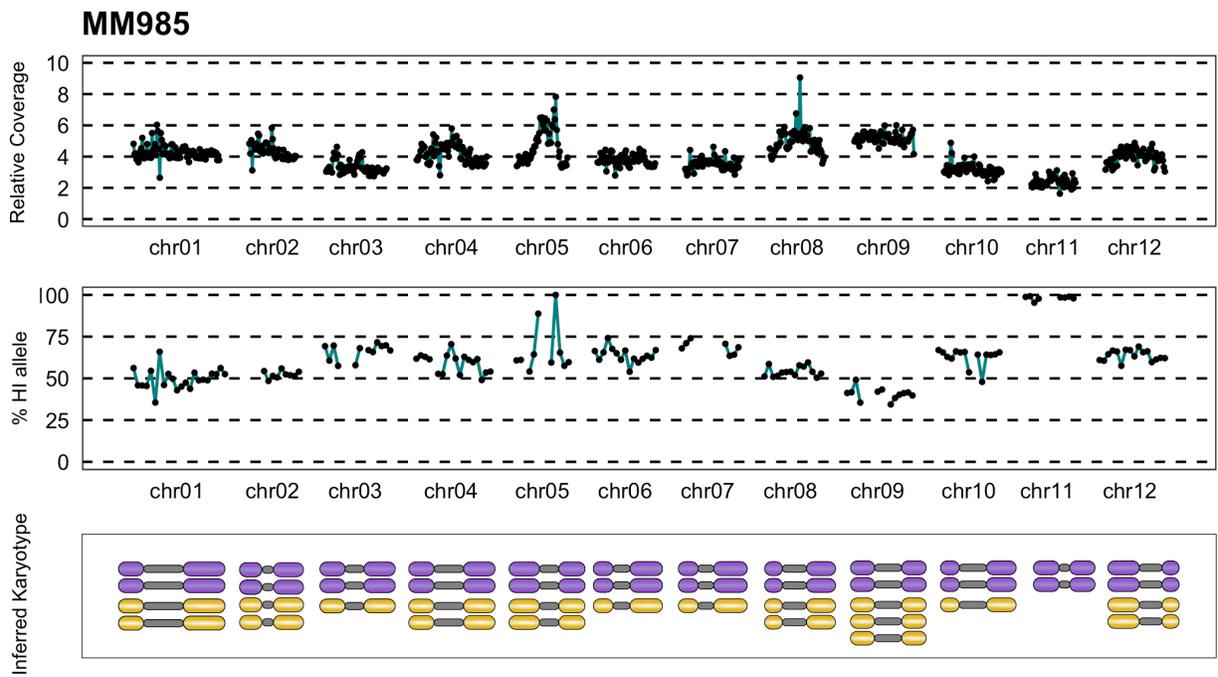
Supplemental Figure S3.6. Dosage variation linkage matrix of WA.077 dihaploids. For 231 dihaploids extracted from WA.077, relative coverage values were derived for 1Mb bins and clustered to infer discrete dosage states for each bin. Non-random association between polymorphic bins were assessed by the Fisher Exact test. Blue points: intrachromosomal linkage (False Discovery Rate = 0.05), black points: interchromosomal linkage (FDR = 0.05).



Supplemental Figure S3.7. Test of potato haploid inducer (HI) haplotype phasing. From the HI chromosome 11 haplotype extracted from HI addition line MM246 (CIP315048.40), only loci where the phased allele did not match the tetraploid allele were retained. Each panel corresponds to one triploid hybrid. For each hybrid, the percentage of the phased, retained allele (H' allele) was across SNP loci in non-overlapping 4Mb bins of the reference genome and plotted. The non-recombining region of chromosome 11 reported in (Bourke et al., 2015) is shaded in gray. If the triploid hybrids inherited the same centromere 11 HI haplotype as MM246, approximately 33% H' allele is expected throughout the entire centromere. If the triploid hybrid inherits the HI centromere 11 haplotype that was not observed in MM246, 0% H' allele is expected.



Supplemental Figure S3.8. Haploid Inducer (HI) centromeric heterozygosity for tetraploid hybrids of the potato haploid induction cross. Phased HI haplotypes were filtered to retain only loci where the phased HI allele and tetraploid allele were the same, yielding a complementary SNP marker set to the results shown in Fig. 3.5C. Each point corresponds to the percentage of the phased (H') allele among reads spanning the non-recombining regions (coordinates from (Bourke et al., 2015)) of a chromosome of one tetraploid hybrid. Chromosome 8 was used to assess IVP101 hybrids. Chromosome 10 was used to assess PL4 hybrids. Chromosome 11 was used to assess IVP35 hybrids. Point shapes indicate whether the assayed chromosome was euploid or euploid.



Supplemental Figure S3.9. Dosage plot, SNP plot and inferred karyotype of WA.077 x IVP35 chromosome 11 disomic tetraploid potato hybrid.

Supplemental Tables

Supplemental Table S3.1. Genomic content and embryo spot phenotype in progeny of potato haploid induction crosses. For each progeny, ploidy was estimated from flow cytometry or chloroplast counting and presence or absence of the dominant and haploid inducer-specific embryo spot phenotype (Hermsen and Verdenius, 1973) was recorded.

Ploidy	Without embryo spot	With embryo spot
2x	917	2
3x	4	26
4x	3	101

Supplemental Table S3.2. Description of 19 tetraploid potato clones used for dihaploid extraction.

Accession Name	CIP Accession Number	Category
458	CIP391931.1	Advanced clone
93.003	CIP390637.1	Advanced clone
Atlantic	CIP800827	Variety
C01.020	CIP301023.15	Advanced clone
Tacna (C90.170)	CIP390478.9	Variety
C91.640	CIP388615.22	Advance clone
C92.172	CIP392780.1	Advanced clone
C93.154	CIP392820.1	Advance clone
Desiree	CIP800048	Variety
LR00.014	CIP300056.33	Advanced clone
LR00.022	CIP300072.1	Advanced clone
LR00.026	CIP300093.14	Advance clone
LR-93.073	CIP392822.3	Advanced clone
LRY-3.57	CIP313047.57	Advanced clone
LRY-21.25	CIP313065.25	Advanced clone
WA.073	CIP397099.4	Advanced clone
WA.077	CIP397077.16	Advance clone
WA.104	CIP397073.16	Advanced clone
Maria Bonita (Y84.027)	CIP388676.1	Variety

Supplemental Data Sets

Supplemental Data Set S3.1. IGV Screenshots of short read alignments to v4.04 reference genome showing ambiguous introgression junctions.

Supplemental Data Set S3.2. Description of plant material used in this study

Supplemental Data Set S3.3. Summary of sequencing libraries constructed or analyzed in this study

All Supplemental Data Sets are available on Dryad at
<https://datadryad.org/stash/dataset/doi:10.25338/B8JS8D>

Chapter 4

Genetic mechanisms underlying somatic chromosomal changes in potato

[Unpublished]

Kirk R Amundson¹, Benny Ordoñez^{1,2}, Xin Zhao^{1,3}, Guilherme T Braz^{4,5}, Michelle Fossi¹, Isabelle M Henry¹, Jiming Jiang^{4,6} and Luca Comai^{1*}

¹ Plant Biology and Genome Center University of California, Davis, 1 Shields Avenue, Davis, CA, 95616, USA.

² Integrative Genetics and Genomics Graduate Group, University of California, Davis, CA 95616, USA.

³ Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China.

⁴ Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA.

⁵ Current address: Biology Institute, State University of Campinas, Campinas, Brazil.

⁶ Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA.

ORCID IDs: 0000-0003-3435-626X (K.R.A.); 0000-0003-4554-4354 (B.O.); 0000-0002-0803-7778 (G.T.B.); 0000-0002-6796-1119 (I.M.H.); 0000-0002-6435-6140 (J.J.); 0000-0003-2642-6619 (L.C.)

*Corresponding author: lcomai@ucdavis.edu

Short title: Somatic mutations in regenerated potatoes

The author responsible for the distribution of materials integral to these findings presented in the article in accordance with the policy described in the Instructions for Authors is: Luca Comai (lcomai@ucdavis.edu)

Abstract

Regeneration of plants from callus frequently results in unpredictable, and often undesirable, phenotypic change, a phenomenon known as somaclonal variation. These changes can be caused by changes to chromosome number and structure. To evaluate the contribution of altered chromosome structure, we analyzed potatoes regenerated from leaf protoplasts for changes in copy number and heterozygosity. Individual instances of chromosome-wide copy-neutral change of heterozygosity, and complex rearrangements resembling chromoanagenesis were observed. Recurring copy number and change of heterozygosity variation was also observed in multiple clones. Detailed investigation of the origin of this variation determined that it was due to mosaicism of an unbalanced chromosome translocation that was present in the original protoplast donor. We hypothesize that these changes spontaneously arose during somatic growth and became fixed in shoot apical meristem layers L2 and L3 of the protoplast donor. Protoplast regenerants that originated from the L2 or L3 layer were therefore genetically distinct from the others. These findings indicate that chromosome rearrangements incurred either before or during tissue culture regeneration can contribute significantly to somaclonal variation.

Introduction

Plants have the remarkable ability to regenerate from differentiated cells. This property requires specialized tissue culture conditions. It is a cornerstone of modern plant biology, as nearly all transgenic and gene-edited plants require the regeneration of individual cells that carry a genetic modification of interest or the potential to make one. Despite over 50 years of technological advances in plant tissue culture, regeneration remains a severe bottleneck that hinders advances in genome editing technology (Yuan et al., 2021; Li et al., 2020; Lin et al., 2020; Tang et al.,

2020; Decaestecker et al., 2019; Wolter et al., 2018; Čermák et al., 2017; Baltes et al., 2014), delivery of editing reagents into plant cells (Maher et al., 2019; Kelliher et al., 2019; Demirer et al., 2019) and high-throughput identification of editing targets (Kim et al., 2021; Lozano et al., 2021; Ramstein and Buckler, 2021; Zhang et al., 2019; Johnsson et al., 2019; Ramu et al., 2017). This bottleneck exists for several reasons. For example, different species, as well as genotypes within species, often require empirical optimization of regeneration conditions. Regeneration is both technically demanding and labor-intensive, and as a result, transformation of several important species such as maize is now carried out by a few specialized laboratories (Altpeter et al., 2016). Furthermore, callus regeneration is mutagenic in virtually every plant species studied (Neelakandan and Wang, 2012; Veilleux and Johnson, 2010; Phillips et al., 1994; Larkin and Scowcroft, 1981).

Extensive studies of plants regenerated or maintained in tissue culture have revealed changes to DNA methylation (Han et al., 2018; Stelpflug et al., 2014; Stroud et al., 2013; Vining et al., 2013; Kaeppler and Phillips, 1993b, 1993a), transposable elements (Miyao et al., 2012; Sato et al., 2011b; Kikuchi et al., 2003; Hirochika et al., 1996; Peschke and Phillips, 1991; Brettell and Dennis, 1991; Peschke et al., 1987), DNA sequence (Park et al., 2020; Tang et al., 2018; Zhang et al., 2014; Miyao et al., 2012; Jiang et al., 2011), chromosome number and chromosome structure (Karp et al., 1987; Wheeler et al., 1985; Ramulu et al., 1983; Karp et al., 1982; Pucker et al., 2019; Fossi et al., 2019; Gernand et al., 2007; Gill et al., 1987; Lee and Phillips, 1988). It is generally accepted that most of the observed variation is induced by the regeneration process, as evidenced by the onset of chromosomal abnormalities in callus (Sree Ramulu et al., 1984), genetic heterogeneity among plants regenerated from the same callus (Fossi et al., 2019; Ramulu

et al., 1983; Thomas et al., 1982; Karp et al., 1982), a trend toward more frequent and severe genome instability with increased duration of callus phase (Hirochika et al., 1996), and RT-PCR quantification of TE excision events before and after tissue culture (Sato et al., 2011a).

While somatic DNA and phenotypic changes are frequent during tissue culture, plants can produce somatic variants during normal growth as well, a process called sporting (Foster and Aranzana, 2018). Sports are important in breeding of clonally propagated species, such as fruit trees (Tan et al., 2019; Wang et al., 2021; Guan et al., 2021), grapes (Carbonell-Bejerano et al., 2017; Pelsy et al., 2015), and potatoes (Bethke et al., 2014; Miller et al., 1999; Leever et al., 1994). Notably, in potatoes, the adoption of improved spontaneous sports is common, as exemplified by a series of Russet Norkotah somatic mutants (Miller et al., 1999). How do changes induced in tissue culture relate to those that produce sports during normal growth? While tissue culture is artificial, transcriptional analysis indicates that it shares cellular and molecular pathways with wound healing, as well as stress responses (Florentin et al., 2013; Ikeuchi et al., 2013; Fehér, 2015; Grafi and Barak, 2015) that can be accompanied by widespread TE activation (Hirochika et al., 1996; Miyao et al., 2012).

Potato (*Solanum tuberosum* L.) is an attractive system for studying somaclonal variation due to its autopolyploid genome, relative ease of regeneration from protoplasts and explants, and vegetative propagation through tubers or *in vitro* cuttings. Cytological studies, and more recently, low-coverage whole genome sequencing, have identified frequent changes to chromosome number and structure among potatoes regenerated from protoplasts or explants (Fossi et al., 2019; Gill et al., 1987; Fish and Karp, 1986; Sref Ramulu et al., 1986; Wheeler et

al., 1985; Ramulu et al., 1983; Karp et al., 1982; Creissen and Karp, 1985). While somaclonal variation can confer desirable phenotypes in potato (Tegg et al., 2013; Wilson et al., 2010; Thieme and Griess, 2005; Sebastiani et al., 1994; Matern et al., 1978; Secor and Shepard, 1981; Shepard et al., 1980), improved somaclonal variants without accompanying negative attributes are rare. Potato chromosomes are difficult to distinguish by morphology alone (Dong et al., 2000), but low-coverage sequencing is sufficient for detecting aneuploidy and large structural variants (Amundson et al., 2020b). With inexpensive whole genome sequencing, improved genome assemblies (Sun et al., 2021; Pham et al., 2020; Zhou et al., 2020; van Lieshout et al., 2020), established genetic resources such as haploid inducers (Ordoñez et al., 2021; Hermsen and Verdenius, 1973), genotype data (Prodhomme et al., 2020; Sharma et al., 2018; Pham et al., 2017; Hardigan et al., 2017; Hirsch et al., 2013) and recently developed cytogenetic tools (Braz et al., 2018; He et al., 2018), potato could provide more detailed insight into the mechanistic basis of somatic chromosomal change. For example, regeneration-induced variation, preexisting inherited variation and somatic mutations can be distinguished from one another, and the underlying mechanisms of each can be inferred.

In our previous study (Fossi et al., 2019), we identified recurring structural variants and karyotypic mosaicism among 52 plants regenerated from protoplasts and explants, including widespread aneuploidy, all of which were inferred from low coverage sequencing (Fossi et al., 2019). The current study was designed to better understand the underlying basis of somatic mutations associated with tissue culture regeneration in potato. A better understanding of the structure and patterns of these variants should provide insight into the underlying mechanisms by which somatic variation emerges in clonal polyploid plants.

Results

To further characterize the genomic changes associated with regeneration, we sequenced the genomes of the protoplast-regenerated clones reported by Fossi et al. (2019), the protoplast donor clone (PI310467, originally presumed to be cv. Desiree), and an independent clone of cv. Desiree obtained from Cornell University, to 12-90x coverage with short read sequencing (Supplemental Data Set S4.1). Additionally, we generated and skim-sequenced three segregating populations derived from PI310467: primary dihaploids (n=81), S1 progeny of the original accession (n=83) (Supplemental Fig. S4.1), and S1 progeny of its protoplast-regenerated line MF93 (n=47). Publicly available sequence reads and SNP array genotypes were obtained from previous studies (Hardigan et al., 2016, 2017; Pham et al., 2017; Fossi et al., 2019; Amundson et al., 2020b, 2020a; Hirsch et al., 2013; Sharma et al., 2018; Prodhomme et al., 2020) and incorporated in our analyses. Sequence reads were aligned to the updated DM1-3 v6.1 genome assembly (Pham et al., 2020) to call and genotype SNVs and CNVs, which were then used to infer the basis of any chromosomal mutations sustained during vegetative propagation or tissue culture regeneration.

PI310467 displayed an unbalanced chromosome translocation

First, we characterized the protoplast donor, an accession held at the US Potato Genebank (USPG) as PI310467 and listed as cv. Desiree. Sequence read depth analysis revealed large CNV on the long arm ends of chromosomes 7 and 8. Specifically, PI310467 carried five copies of the terminal 5.6 Mb of chr07, and three copies of the terminal 4.6 Mb of chr08 (Fig. 4.1A). Primary dihaploids exhibited recurring duplication and deletion states relative to PI310467 at both chromosome ends of the respective long arms, indicating that both CNVs were heritable and

segregated as intact haplotypes (Supplemental Fig. S4.2). Within each dosage variable region, the up- and down-haplotypes segregated 1:1 ($\chi^2_{1:1}=2.778$, $df=1$, $p=0.1$), consistent with tetrasomic inheritance of a simplex locus by a gametic population. Furthermore, dosage states between chromosomes were in complete linkage disequilibrium (LD) with each other: without exception, gain of the chr07 end was associated with loss of the chr08 end, and loss on chr07 with gain on chr08 (Supplemental Fig. S4.2), suggesting that the additional chr07 segment is physically linked to the chr08 homolog with the deletion. Consistent with this arrangement, the S1 population showed 3 dosage states (up, middle, and down) of the same chr07 and chr08 regions. The dosage states in each region fit a 1:2:1 ratio ($\chi^2_{1:2:1}=3.892$, $df=2$, $p=0.14$), and dosage states between chromosomes were also in complete LD with each other (Supplemental Fig. S4.3). Hybridization of chromosome-specific oligonucleotide probes to root nuclei revealed that one of the four chr08 homologs carried a chr07 segment at its tip (Fig. 4.1A), confirming an unbalanced translocation, tr8-7. To determine if the translocation occurred after transfer of PI310467 to UC Davis, we acquired and skim sequenced three independent PI310467 cuttings from the USPG stock center, and found that they also carried the tr8-7 CNVs (Supplemental Fig. S4.4). We concluded that PI310467 carries an unbalanced translocation, in which the 5.6 Mb terminal segment from the long arm chr07 substituted the 4.6 Mb terminal segment from the long arm of chr08.

Somatic loss of unbalanced translocation observed among regenerants

The PI310467 regenerants were polymorphic for tr8-7: by read depth analysis, six of them displayed CNV consistent with its absence, suggesting loss of tr8-7 (Fig. 4.1B, Supplemental Fig. S4.5). Chromosome painting chr07 and chr08 in MF93 root cells revealed loss of the fusion chromosome and four apparently normal chr07 and chr08 copies, confirming the loss of tr8-7

(Fig. 4.1B). To test whether the regenerants were true clones of PI310467, we carried out a phylogenetic analysis of 1,019 potato lines, including PI310467 and all regenerants, using 1,858 SNP loci. All regenerants displayed $\geq 93\%$ average pairwise Identity By State (IBS) typical of replicated clones (Fig. 4.1C, Supplemental Fig. S4.6, Supplemental Data Set S4.2), which ruled out strain contamination. Although PI310467 and the regenerants appeared to be true clones of each other, they were neither clones of Desiree (Fig. 4.1C) nor of any other accession tested (Supplemental Data Set S4.2). PI310467 did not appear to be either parent of Desiree either. One parent, Urgenta, was included in the panel and showed 78% IBS with PI310467. Genotype data of the other parent, cv. Depesche, was not available, but Depesche exhibits white-yellow tuber skin and short height uncharacteristic of PI310467

(<https://www.europotato.org/varieties/view/Depesche-E>). The closest relatives were other Desiree clones at 85% IBS. To test whether PI310467 was a Desiree self, we looked for PI310467 heterozygosity at 1.07 million high-confidence Desiree-homozygous loci distributed throughout the genome. PI310467 was heterozygous at 88.3% of tested loci (Supplemental Fig. S4.7), indicating that it is not a Desiree self. While we could not determine the identity of PI310467, these results indicate that it is closely related to Desiree, and that the chromosomal variations observed among protoplast regenerants, including the recurring loss of tr8-7, were somatic in origin.

Mosaicism of PI310467 resulted in two genetically distinct regenerant types

Next, we looked for genomic evidence of known mechanisms that could explain a somatic loss of tr8-7 in the protoplast regenerants. Three possible explanations are i) somatic chromosome substitution (Robinson, 2000), (ii) break induced replication (BIR) (Pâques and Haber, 1999), and (iii) mitotic crossover (Carlson, 1974). The first two are well described in yeast and animals,

but clear examples in plants have not been described (but see (Schubert et al., 2011) for an example of BIR-like conservative DNA replication in plants). In a diploid, all result in loss of heterozygosity (LOH): for a substitution, across an entire chromosome, and for BIR or somatic CO, from a breakpoint to the telomere. In heterozygous polyploids, these same events result in LOH, or, depending on the haplotypes involved, more complex changes to allele dosage, i.e., change of heterozygosity (COH). We did not detect COH of the tr8-7 proximal region for 4 of the 6 somaclones that lacked tr8-7, suggesting that tr8-7 loss was not due to chromosome substitution (Fig. 4.2A). The two somaclones that did show both tr8-7 loss and tr8-7-proximal COH were chr08 trisomics (Supplemental Fig. S4.5). Unexpectedly, all six regenerants that lacked tr8-7 also showed putatively novel alleles relative to PI310467 at 3,163 SNP loci distal to the tr8-7 breakpoint, while regenerants that retained tr8-7 did not (Fig. 4.2B). Apparent *de novo* mutations along the recovered chromosome segment suggests that tr8-7 loss was not due to a somatic crossover between tr8-7-carrying and tr8-7-lacking chr08 homologs. Regenerants lacking tr8-7 showed the same alleles, despite originating from independent calli (Fig. 4.2C). Further inspection revealed that these SNPs were present in PI310467 at approximately 5% variant allele frequency (VAF) (Fig. 4.2D). This suggests that PI310467 is a mosaic of two cell lineages: one rare lineage that carries both a balanced karyotype (with respect to chr07 and chr08) and the SNPs on chr08, and a more abundant lineage that carries tr8-7. Hereafter, we define these 3,163 low-VAF SNP alleles that preexisted in PI310467 as “uncovered SNPs”. They were detected at higher frequency in the regenerants that lacked tr8-7.

Most of the sampled leaf cells carried tr8-7 (Fig. 4.1A). Furthermore, tr8-7 was detected in L3-derived adventitious roots (Fig. 4.1B) and transmitted through the germline, indicating its

presence in L2 (Supplemental Fig. S4.2, S4.3). Based on these results, we hypothesized that PI310467 is a periclinal mosaic, with L2 and L3 layers composed of cells that carry tr8-7, and L1 made up of cells without tr8-7. If so, the uncovered SNPs are not expected to be transmitted to PI310467 primary dihaploids and S1 progeny. Low coverage sequencing of all 289 progeny was pooled, and despite 37x coverage of the pool, the uncovered SNPs were not detected (not shown). This could be explained by the exclusive presence of tr8-7 cells without tr8-7 in the L1 layer, or alternatively, by sterility or lethality of the associated chr08 haplotype. To test these two hypotheses, cell layer-specific gDNA fractions of PI310467 could be genotyped for the predicted layer-specific SNPs. Additionally, skim sequencing the MF93 S1 population could determine whether the alleles associated with tr8-7 loss are germline transmissible. Both experiments are in progress, but are not included in this thesis for timing reasons.

To determine whether the uncovered SNPs were PI310467 mutations, we genotyped a panel of 99 clones representing potato diversity. This panel included PI310467 and all protoplast regenerants as well as tetraploid cultivars, landrace clones, and diploid *Solanum* species with publicly available whole genome sequence data. As the panel includes multiple regenerants that carry the uncovered SNPs, the expected allele frequency of a PI310467 mutation, which would appear in only PI310467 and regenerants lacking tr8-7, was 3.48%. The observed frequencies of these alleles in the panel ranged from 0.3% to 54%, and 94% of the alleles were observed at frequencies exceeding 3.48% (Fig. 4.3A). This indicated that the uncovered SNPs were also present in other potatoes, and were neither PI310467-specific mutations nor read mapping artifacts (Wijnker et al., 2013). The uncovered SNPs were detected in simplex or higher allele dosage in cultivars released as early as 1857 (Garnet Chili), in landrace clones, and in wild

Solanum species (Fig. 4.3B). Read support was robust, with higher dosage of the uncovered SNPs compensating for lower sequencing coverage in some cases (Fig. 4.3C). These results demonstrate that the uncovered SNPs are both common and ancestral. Regarding tr8-7, these results show that it was the derived state within PI310467, and that most of the soma and entire germline carry it.

Chromoanagenesis

Previously, we documented instances of complex chromosomal changes among protoplast regenerants (Fossi et al., 2019), resembling the type of catastrophic genome restructuring first documented in cancer genomes: chromothripsis or chromoanasythesis. Both types of events are characterized by clusters of dosage variation within a single chromosome, or within a segment of a single chromosome (Pellestor, 2019; Holland and Cleveland, 2012). Chromothripsis is the NHEJ-dependent restructuring of a chromosome. In a diploid, it mostly involves oscillations between copy numbers of 1-3 (Korbel and Campbell, 2013; Stephens et al., 2011). In contrast, chromoanasythesis occurs when replication forks stall, causing DNA breaks, and DNA ends undergo replicative repair during which they switch templates. Iterative duplications can then result in higher copy numbers of certain chromosome regions coupled to complex rearrangements of the duplicated and intervening regions (Pellestor and Gatinois, 2018). To infer the mechanism that resulted in complex chromosomal rearrangements among somaclones, we performed detailed genomic and cytogenetic characterization of MF74, a regenerated clone that displayed unusual copy number variation on the right arm of chr08 (Fossi et al., 2019).

To determine how these CNVs were physically arranged, we carried out oligo FISH using chromosome 7- and chromosome 8-specific probes, as described above. One of the chromosome

8 homologs underwent an expansion in one of the arms, and a second homolog was both truncated, and carried a region that was unlabeled by either chr07 or chr08-specific probes (Fig. 4.4A). This region could correspond to centromeric repeats of chr08, which were intentionally excluded from the chromosome-specific probe sets (He et al., 2018), or to a segment of a chromosome that was not labeled. To corroborate these findings, we also carried out oligo-FISH with a probe library that labeled each of the 12 chromosomes with a unique dual-color barcoding pattern (Braz et al., 2018). The probe libraries were expected to label the long arm of chromosome 8 with two green foci. Two chr08 homologs showed the expected pattern, but green foci were not detected on the truncated chr08 homolog, indicating that the long arm had been lost (Fig. 4.4B). In contrast, the expanded arm showed additional FISH foci, suggesting amplification of one region tiling the probe set, or alternatively, translocation between chr08 homologs. We also detected a second, copy neutral but complex chromosome rearrangement involving unknown chromosomes (Fig. 4.4B, arrow).

We then carried out read depth-based CNV analyses of PI310467 and MF74. Copy number variants included pentasomy of chr04, partial trisomy of chr02, and the complex CNVs of chr08 (Fig. 4.4C, Supplemental Fig. S4.8). Regarding the partial trisomy of chr02, the distal 2 Mb of the long arm was present in 4 copies, while the rest the chromosome arm was present in 3 copies and the short arm, which carries the rDNA arrays, was present in 1 copy in both PI301467 and MF74 (Fig 4.4D). On chr08, in PI310467, copy number oscillated between 0, 1, 2, and 4, but not 3 throughout the pericentromere (Fig. 4.3D). MF74 CNV patterns were similar through the left euchromatic arm and into the pericentromeric heterochromatin until approximately 20 Mb. At that point, MF74 copy number oscillated between 2 and 3 through the centromere into the right

pericentromeric heterochromatin. Chromosome segments along the right arm were present in 1-6 copies (Fig. 4.4D). The local amplification of chromosome segments along the right arm detected by genome sequencing and FISH, from 3 copies in PI310467 to 5-6 copies of some regions in MF74, suggests chromoanasythesis. None of the other protoplast regenerants exhibited this pattern (Supplemental Fig. S4.5). Although this was the only instance of chromoanasythesis-like rearrangements in our panel, it provided additional evidence of the types of complex chromosome rearrangements that can be sustained by tissue culture regenerated plants.

Chromosome substitution

Uniparental disomy can occur when chromosome missegregation results in a trisomy or monosomy, followed by a return to euploidy by subsequent missegregation of another chromosome. This is expected to result in a chromosome-wide LOH, or in a polyploid, in COH as described above. We looked for these changes by comparing allele-specific read depth between PI310467 and the regenerants. To provide an example of the type of signal expected from such a change, we used the haplotype-resolved assembly of cv. Otava (Sun et al., 2021) to simulate short reads of a whole-chromosome haplotype exchange. We then aligned simulated reads to DM1-3 v6.1 and compared allele dosage at Otava SNPs. By plotting the fraction of reads with the reference allele by the position of the reference genome, clusters of SNPs with the same allele dosage were observed across the chromosome (Supplemental Fig. S4.9A). As expected, the simulated chromosome substitution produces a chromosome-wide shift in allele dosage at heterozygous loci (Supplemental Fig. S4.9B). To compare allele dosage at the same loci, we then partitioned the loci by their genotype in Otava, and then, for each locus, plotted the difference in reference allele proportion between Otava and the simulated chr01 substitution line

(Supplemental Fig. S4.9C). Allele dosage was altered at some, but not all loci, consistent with a change in dosage when the lost and duplicated haplotypes were polymorphic at a locus.

Using this approach, we detected allele dosage clustering along chr01 in PI310467 and regenerant MF84 (Fig. 4.5A). Similar to the simulated Otava chromosome substitution, a chromosome-wide shift in allele dosage was observed between PI310467 and MF84 (Fig. 4.5B). By read depth, we did not detect a net change in copy number along the gene-rich parts of the chromosome arms of PI310467 and MF84, though CNVs could be observed throughout the gene-poor pericentromeres (Fig. 4.5C). This can be explained by the existence of large structural variation, mainly indels, in the pericentromere haplotypes. Therefore, the chromosome-wide COH observed in MF84 is consistent with loss of one chr01 haplotype and duplication of another. Although a single instance in our panel of 12 regenerants, these data indicate that compensating chromosomal missegregation sustained during tissue culture can change heterozygosity without a net change to chromosome copy number.

Discussion

Our analysis of genomic changes in tetraploid potatoes regenerated from leaf protoplasts revealed a variety of changes in just a limited number of regenerants (n=52). These fall into three categories: chromoanagenesis, chromosome substitution and chromosome translocation. The first two types were detected in individual regenerants, indicating that they likely arose during regeneration, while the latter pre-existed in the protoplast donor, PI310467. These examples, together with widespread aneuploidy (Fossi et al., 2019), and simple chromosome break and

repair, provide examples of the genomic changes that contribute to somaclonal variation in tissue culture, as well as somatic genome instability under normal growth conditions.

Chromosome substitution

A potentially deleterious outcome of genome instability is loss of heterozygosity (LOH), which has contributed to notable horticultural sports (Tan et al., 2019; Carbonell-Bejerano et al., 2017; Pelsy et al., 2015). Known causes of LOH include deletions, gene conversion, mitotic recombination, and break-induced replication. Among our panel of 12 regenerants, we identified one instance of altered allele dosage across an entire chromosome with no net change to chromosome copy number (Fig. 4.5). This change of heterozygosity (COH) event resulted in altered allele dosage in a haplotype-dependent manner: LOH is possible but not guaranteed, and dosage of a particular allele may increase. Given the widespread aneuploidy observed among callus regenerants, the detected COH events can likely be explained by compensating chromosome missegregation events. This may explain primary regenerants of wheat observed to be true-breeding for tissue culture-induced mutations (Larkin et al., 1984). In humans, somatic chromosome missegregation can result in uniparental disomy, and chromosome-wide LOH that can be associated with either Prader-Willi or Angelman syndrome (Fridman and Koiffmann, 2000; Mascari et al., 1992). Little is known about the extent and impact of somatic chromosome substitution in plants. Further study, especially in polyploids with increased buffering against deleterious mutations and dosage imbalance, could shed light on the frequency and impact of somatic chromosome substitution in plants.

Chromoanagenesis

Chromoanagenesis was originally described in cancer genomes (Stephens et al., 2011) and has since been demonstrated in progeny of CENH3-mediated haploid induction of *Arabidopsis* (Tan et al., 2015) and poplar (*Populus spp.*) trees produced from gamma-irradiated pollen (Guo et al., 2021). In these cases, genomic evidence of copy number increases was critical for mechanistic inference. Recently, a single anaphase bridge was shown to be sufficient to trigger a cascade of genome instability outcomes, including chromoanagenesis (Umbreit et al., 2020). Cytological events associated with chromoanagenesis, such as anaphase chromatin bridges, micronuclei, and complex rearrangements have been observed, for example, in *Allium fistulosum* callus cells (2007). We still do not know when and how plant genomes are destabilized in tissue culture, but a fair assumption is that instability is triggered rapidly and can persist after regeneration. The buffering of deleterious variation and the ability to clone potatoes through *in vitro* cuttings or tubers could be advantageous for preserving chromoanagenesis outcomes for further evaluation and study. For practical applications, these extreme chromosome rearrangements are probably undesirable. The chromosome rearrangements of MF74 resulted in a severe vegetative growth deficiency that could have been culled early, but this need not have been the case. Somatic mutations could have been associated with reproductive deficiencies. Avoiding the often deleterious mutations associated with somaclonal variation will require transformation and genome editing protocols that circumvent tissue culture (Nasti and Voytas, 2021; Altpeter et al., 2016).

PI310467 is not Desiree

The phenotypes of PI310467, which is held as cv Desiree at the US Potato Genebank, resembles Desiree in many respects: late maturity, good self fertility, and pink-skinned tubers with yellow flesh. However, our genomic analyses show that PI310467 is not a clone of Desiree, nor of any of 1,000 potatoes we analyzed, nor a parent or selfed offspring of Desiree. Sequencing and analysis of replicated samples traced the identity of PI310467 as an unknown clone back to the US Potato Genebank to some time before 2015. We caution that institutions working with Desiree sourced from the gene bank after this time are unlikely to have Desiree. It is possible, but less parsimonious, that PI310467 is the true Desiree. In ongoing work, we are evaluating PI310467 and Desiree obtained from Cornell University for diagnostic traits.

Unexpected mosaicism of PI310467

Changes observed in tissue culture may be related to those that produce sports during somatic growth. While tissue culture regeneration is artificial, culture-derived instability may accelerate rare somatic mutations that are comparable to those occurring during normal vegetative growth. Yet another possibility is that somaclonal variants reflect diversity preexisting in the source plant, either recently arisen, such as *de novo* variants formed during leaf development, or long standing, such as a periclinal chimera. Here, we found that PI310467 is a genetic mosaic of two cell types, one of which carries a spontaneous translocation, which explains the recurring structural variation reported previously for chr07 and chr08 (Fig. 4.6) (Fossi et al., 2019). The translocation was detected in both L2-derived offspring and L3-derived adventitious roots of PI310467. SNP alleles that appeared as simplex in regenerants lacking tr8-7, and were present at substoichiometric VAF in PI310467, were not inherited by its progeny. At the same time, these

SNPs were common among a potato diversity panel and present in older cultivars, landraces, and wild relatives, indicating that PI310467 originally lacked tr8-7. We infer that a somatic rearrangement resulting in tr8-7 arose in a single cell of either the L2 or L3 layer of PI310467 (Fig. 4.6). Tr8-7 could have occurred by a mitotic crossover between ectopic sequences, a large-scale gene conversion event resembling BIR, or nonhomologous end joining. Characterization of the tr8-7 junction sequence, which was not possible with short reads, may provide additional insight. However tr8-7 occurred, the cell carrying it then spread to both the L2 and L3 layers. This would have occurred despite a genomic imbalance relative to surrounding cells, which testifies to the buffering capacity of the tetraploid potato genome. No changes to PI310467 cell composition clones have been detected in over six years of vegetative propagation, which agrees with reports of stable L1 mosaicism in potato (Howard, 1971). In contrast, invasion of L2 cells into the L3 is frequent enough to explain a derived translocation in the L2 and L3 (Howard, 1972; Howard et al., 1964). Genotyping of layer-specific gDNA fractions will be necessary to validate that PI310467 is a periclinal mosaic. The availability of clone MF93, a regenerant that lacks tr8-7 and displays no other chromosomal abnormalities, provides the opportunity to evaluate the phenotypic effect of tr8-7. These experiments are in progress and will be part of a forthcoming publication based on this chapter.

The chromosomal rearrangements described here serve as a framework for investigating genomic changes associated with naturally occurring bud sports. Examples in potato include somatic mutants known as “giant hill” that often are higher yielding, later maturing and can show novel disease resistance. Giant hill varieties are common but the causative mutations are unknown (Miller et al., 1999). Other bud sports of potato show changes in tuber skin color, for

example, Norland, Red Norland and Dark Red Norland are sports with deeper red tuber skin, and similarly, Red La Soda is a sport of La Soda (Miller, 1954). These sports may be an attractive system for studying the phenotypic somatic mutations, as potato genes controlling anthocyanin biosynthesis (Zhang et al., 2009; Jung et al., 2005) and accumulation during development (Laimbeer et al., 2020; Jung et al., 2009) have been cloned. Perhaps the most famous example is Russet Burbank, which arose as a sport of Burbank through an unknown mutation (Bethke et al., 2014). Reversion to the non-russeting Burbank trait is frequent among somaclonal variants of Russet Burbank (Shepard et al., 1980), raising the question of chimerism. Russet Burbank clones regenerated from tissues enriched for one or the other somatic layer, however, displayed comparable frequency of smooth skin, suggesting lack of periclinal mosaicism (Nassar et al., 2008). Compared to our knowledge of color traits, the genetic basis of russeting in potato is less understood; segregation ratios in diploid mapping populations indicate that dominant alleles at three independently segregating loci are required for russeted tuber skin (Jong and De Jong, 1981), but the responsible genes remain unknown. Emerging biochemical, molecular and genetic evidence on suberin regulation may provide gene candidates, such as a MYB transcription factor known to affect russeting in apples (Legay et al., 2016). What can be learned by comparing bud sports of potato with their respective progenitors? Identification of causal mutations in bud sports, some of which are chromosomal rearrangements, has shed light on genes controlling traits of interest in *Citrus* (Wang et al., 2021), peach (Tan et al., 2019) and grapevine (Carbonell-Bejerano et al., 2017), and may do so in potato as well.

Here, we show that chromosomal changes in regenerated potato plants are produced by different mechanisms, including chromosome substitution, chromoanagenesis, and unbalanced

translocation. Further, changes can arise due to genetic mosaicism of the plant that provided the explants for tissue culture. The *de novo* events provide a sampling of the genome instability that can occur during regeneration, and to a lesser extent, during vegetative development. The limited sample size of our panel prevents robust inferences on the frequencies of these events, although they are sufficiently high to be easily detected. More sampling of regenerated plants and known somatic mutants may provide more insights on particular outcomes and reveal new ones.

Materials and Methods

Plant material

PI310467, held as cv. “Desiree” in the USDA Potato Germplasm Introduction Station (Sturgeon Bay, WI) was requested in 2015 and 2019. The protoplast regenerants were derived from the 2015 acquisition of PI301467, as described in a previous study (Fossi et al., 2019). Plant material was propagated by *in vitro* cuttings in controlled growth conditions (16h light 25°C: 8h dark 18°C), and by tubers in controlled greenhouse conditions at the University of California, Davis. Haploid induction crosses were carried out between the 2015 acquisition of PI310467 and IVP48 in the greenhouse. Seeds were extracted from mature fruit and recorded for presence or absence of the inducer-specific and genetically dominant embryo spot marker (Hermsen and Verdenius, 1973) and germinated *in vitro* under the same conditions used for stem cutting propagation. The ploidy of each seedling was established by flow cytometric measurement of DNA content against the maternal and paternal parents as standards, as described below. Trichomes were extracted from PI310467 petioles by immersion in a petri dish containing a minimal volume of 50 mM ethylene glycol-bis (2-aminoethylether)-*N,N,N',N'*-tetraacetic acid (EGTA) in 1x phosphate-buffered saline (PBS) and gently scraping a metal forceps against the petiole to

release trichomes into solution. Up to 20 petioles were processed in the same batch. Trichomes were then collected into 1.5ml tubes, centrifuged at 3,800 RCF for 1 minute and washed three times with 1x PBS. Genomic DNA was extracted from trichomes using a CTAB extraction (Ghislain M., Zhang D. P., Herrera, M. R., 1999) with the trichome pellet in place of leaf tissue.

Flow cytometry

Approximately 50-60 mg of greenhouse-grown leaf tissue was harvested from each sample and homogenized in 500 μ l of LB01 buffer (Doležel et al., 1989) and left to rest for 1 minute. 250 μ l of homogenate was passed through a 20 μ m filter (Partec 04-0042-2315) into tubes containing 12 μ l of 1mg/ml propidium iodide and 2.5 μ l of 5mg/ml RNase. Samples were incubated in the dark for 5 minutes and analyzed in an Accuri C6 flow cytometer (BD biosciences), with the following filter configurations: a) FL-1 530/14-nm bandpass filter, b) FL-2 585-20nm bandpass filter and c) FL-3 670-nm longpass filter. Threshold levels were set to 10,000 for forward scatter (FSC) with a secondary threshold of 1,000 for FL-2 (Galbraith et al., 2011).

Library construction and sequencing

Genomic DNA was extracted from greenhouse-grown plants as previously described (Ghislain M., Zhang D. P., Herrera, M. R., 1999). Sequencing libraries were either prepared in-house or by Novogene, Inc. For in-house sequencing library preparation, approximately 750ng of genomic DNA was used as input with KAPA Hyper Prep Kit (catalog no. KK8504) and reagents were used at half-scale reactions as previously described (Fossi et al., 2019) sequenced on an Illumina NovaSeq 6000 instrument at the University of California San Francisco Center for Advanced Technologies, and demultiplexed using the custom Python script (allprep-12.py) available from <https://github.com/Comai-Lab/allprep>. Sequencing reads available from previous studies (Fossi

et al., 2019; Amundson et al., 2020b; Hardigan et al., 2017; Pham et al., 2017) were retrieved from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) and incorporated in subsequent analyses. Supplemental Data Set S4.1 contains additional information on sequencing libraries that were either generated or used in this study.

Variant calling

Adapter and low quality sequences were trimmed from raw reads using Cutadapt v1.1.5 (Martin, 2011), with only reads ≥ 40 nt in length retained. Trimmed reads were aligned to the DM1-3 v6.1 assembly (Pham et al., 2020) to which we appended DM1-3 chloroplast and mitochondrial sequences available from <http://spuddb.uga.edu/index.shtml>, using BWA mem (v0.7.12r1039) with default parameters. Alignments were further processed to remove PCR duplicates, soft-clip one mate of an overlapping pair, and filter out read pairs with mates aligning to different chromosomes, as previously described (Amundson et al., 2020a). Single-nucleotide and short indel variants were called and genotyped using freebayes (version 1.3.2) (Garrison and Marth, 2012) with minimum mapping quality 41, minimum base quality 20, population priors off, up to 6 alleles considered per variant, and all other parameters at the default setting. To remove low-quality variants, the following hard filters were applied using bcftools (version 1.12): $EPP \leq 30$, $EPPR \leq 30$, $MQM \geq 30$, $MQMR \geq 30$, $RPP \leq 30$, $RPPR \leq 30$, $SAP \leq 30$, $SRP \leq 30$.

We identified putatively L1-specific alleles as follows: Allele-specific read counts of both PI310467 replicates were merged. From the merged read counts, we observed clusters corresponding to the five genotype classes (nulliplex, simplex, duplex, triplex, quadruplex) and an additional low-VAF cluster. The merged PI310467 replicates corresponded to approximately 120x average genome-wide coverage

Chromosome dosage

Dosage analysis was carried out as previously described (Henry et al., 2015; Fossi et al., 2019; Amundson et al., 2020a). To derive standardized coverage values for each sample, mapped read counts in non-overlapping 1Mb bins of the DM1-3 v6.1 pseudomolecules were normalized to PI301467 and multiplied by the sample ploidy inferred from flow cytometry. No correction was made for CNV present in PI310467.

Phylogenetic analysis

Genotype data from a variety of accessions was retrieved from three previous studies (Hirsch et al., 2013; Sharma et al., 2018; Prodhomme et al., 2020). The dataset of Hirsch et al., (2013) consisted of 250 accessions genotyped at 3,763 loci. Genotypes were formatted as base calls with allele dosage resolved for the tetraploid samples in the panel (i.e., “AAAT” for tetraploids and “AT” for diploids). DM1-3 was included in the dataset. The dataset of Sharma et al., (2018) dataset consisted of 341 accessions genotyped at 5,718 loci. The five possible genotype classes were formatted as dosage of “A” and “B” alleles (i.e., AAAA, AAAB, AABB, ABBB, BBBB), without information on the base that corresponded to the “A” and “B” alleles, and DM1-3 not included in the dataset. The dataset of Prodhomme et al., (2020) consisted of 330 accessions genotyped at 10,968 loci. Genotype calls were formatted as the dosage of the minor SNP allele, with 0 indicating homozygosity for the major allele, and 4 indicating homozygosity of the minor allele. DM1-3 was not included in the dataset.

To merge these datasets, the Hirsch dataset was used to define the allele observed in DM1-3 as the reference allele, and the allele not observed in DM1-3 as the alternate allele. Next, the

genotype calls of the Prodhomme dataset were recoded as reference and alternate alleles as follows. Genotype calls of 11 clones represented in both the Hirsch and Prodhomme datasets (Bintje, Defender, Early Rose, Katahdin, Kennebec, Ranger Russet, Russet Burbank, Spunta, Stirling, Umatilla Russet and Yukon Gold) were extracted, and for each clone, at each SNP, we inferred the identities of the reference and alternate alleles by dosage. For example, if the genotype call of Spunta was “CCCG” in the Hirsch dataset and “3” in the Prodhomme dataset, the putative identities of the major and minor alleles of this SNP in the Prodhomme dataset, according to the Spunta genotypes, were C and G, respectively. For each locus, duplex genotype calls (e.g., “AATT”) were withheld from analysis, and the top-ranking definitions of the major and minor alleles among all clones with ≥ 3 supporting clones were used to define the major and minor alleles. Genotype calls of the entire Prodhomme dataset were then recoded in reference/alternate notation according to the consensus allele definitions at each locus. To recode the “A” and “B” alleles of the Sharma dataset, we compared 16 clones genotyped in the Hirsch and Sharma datasets (Atlantic, Bintje, Chieftain, Dark Red Norland, Defender, Kennebec, Ranger Russet, Red Pontiac, Russet Burbank, Russet Norkotah, Sierra Gold, Spunta, Stirling, Superior, Torridon and Umatilla Russet) using the same consensus approach as the Prodhomme dataset conversion. The merged array dataset was then merged with the genotype calls of all sequenced samples, and missing data were imputed as the mean dosage of the locus across the entire panel. The genotype calls are provided as Supplemental Data Set S4.3.

Next, we calculated average pairwise identity by state (IBS) between all sample pairs of the same ploidy in the panel. For diploid-diploid and diploid-tetraploid comparisons, IBS was calculated as:

$$IBS(a, b) = \frac{1}{2n} \sum_{i=1}^n Sim(g_a, g_b)$$

where $Sim(g_a, g_b)$ was 0 if no alleles were shared, a 1 if one allele was shared, and a 2 if two alleles were shared. Pairwise IBS of tetraploids was calculated as:

$$IBS(a, b) = \frac{1}{4n} \sum_{i=1}^n Sim(g_a, g_b)$$

where $Sim(g_a, g_b)$ was 0 if the two genotypes were identical, a 1 if one allele was shared, a 2 if two alleles were shared, 3 if three alleles were shared and 4 if four alleles were shared. For plotting, samples were clustered using `hclust()` with the “complete” agglomeration method in R version 3.6.2.

FISH

Root tips were harvested from greenhouse-grown plants and fixed in 3:1 ethanol:acetic acid and stored at -20°C until staining. At the time of staining, root tips were digested in a solution of 3% cellulase (Yakult Pharmaceutical, Tokyo, Japan), 1.5% pectinase (Plant Media), and 1% pectolyase (Sigma Chemical, St. Louis, MO) at 37°C for 50 minutes. After digestion, root tips were placed on a microscope slide and macerated with a needle in 20µl of 45% acetic acid. The suspension was then spread with a needle on a hot plate at 50°C for 2 minutes. Chromosomes were fixed by adding 200µl of 3:1 ethanol:acetic acid on a hot plate at 50°C for 10 seconds. Afterward, an additional 200µl of 3:1 ethanol:acetic acid fixative was dropped on the tilted slide and dried at room temperature. Slides were also prepared using the dropping method (Kato et al., 2004) for chromosome painting experiments. FISH was performed as previously described (Dong et al., 2000). The hybridization mixture (500 ng of each labeled probe as ssDNA, 50% formamide, 10% dextran sulfate, 2x SSC) was applied directly to denatured chromosome slides

and incubated for 2 days at 37°C. Approximately 2000 ng of sheared genomic DNA with average size 100bp prepared from *S. etuberosum* and *S. caripense* was used as blocking DNA in chromosome painting experiments. The hybridization mixture for chromosome painting was denatured at 95°C for 8 minutes and incubated at 37°C for 2 hours before being applied to denatured chromosome slides. Biotin- and digoxigenin-labeled probes were detected by anti-biotin fluorescein (Vector Laboratories, Burlingame, CA) and anti-digoxigenin rhodamine (Roche Diagnostics, Indianapolis, IN), respectively. Chromosomes were counterstained with DAPI in VectaShield antifade solution (Vector Laboratories). FISH images were captured using a QImaging Retiga EXi fast 1394 CCD camera attached to an Olympus BX51 epifluorescence microscope. Images were processed with Meta Imaging Series 7.5 software. The final contrast of the images was processed using Adobe Photoshop CS3 software.

Simulations

The publicly available Otava haplotypes (Sun et al., 2021) were retrieved from http://spuddb.uga.edu/otava_potato_download.shtml. Simulated short reads were generated from the Otava haplotypes and from *in silico* constructed chromosome substitution lines with ReSeq (Schmeing and Robinson, 2021). Simulated short read datasets were aligned to the DM1-3 v6.1 assembly, and alignments were processed as described above.

Data Availability

Analysis code is available from <https://github.com/kramundson/ceveza>. Raw sequence reads will be deposited to the National Center for Biotechnology Information (NCBI) Sequence Read Archive upon posting of this manuscript to BioRxiv.

Acknowledgements

This work was supported by the National Science Foundation Plant Genome Integrative Organismal Systems (IOS) Grant 1956429 (Variants and Recombinants without Meiosis) to L.C.

Author Contributions

K.R.A., I.M.H and L.C. conceived experiments. K.R.A., B.O., I.M.H. and L.C. designed experiments, K.R.A., B.O., X.Z. and G.T.B. performed experiments, K.R.A., B.O., G.T.B. and L.C. performed data analysis, B.O., G.T.B and M.F. contributed reagents and materials, and K.R.A., I.M.H. and L.C. wrote the paper with input from all authors.

Conflict of Interest Statement

The authors declare no conflict of interest in this study.

Literature Cited

- Altpeter, F. et al.** (2016). Advancing Crop Transformation in the Era of Genome Editing. *Plant Cell* **28**: 1510–1520.
- Amundson, K.R., Ordoñez, B., Santayana, M., Nganga, M.L., Henry, I.M., Bonierbale, M., Khan, A., Tan, E.H., and Comai, L.** (2020a). Rare instances of haploid inducer DNA in potato dihaploids and ploidy-dependent genome instability. *BioRxiv*.
- Amundson, K.R., Ordoñez, B., Santayana, M., Tan, E.H., Henry, I.M., Mihovilovich, E., Bonierbale, M., and Comai, L.** (2020b). Genomic Outcomes of Haploid Induction Crosses in Potato (*Solanum tuberosum* L.). *Genetics* **214**: 369–380.
- Baltes, N.J., Gil-Humanes, J., Cermak, T., Atkins, P.A., and Voytas, D.F.** (2014). DNA replicons for plant genome engineering. *Plant Cell* **26**: 151–163.
- Bethke, P.C., Nassar, A.M.K., Kubow, S., Leclerc, Y.N., Li, X.-Q., Haroon, M., Molen, T., Bamberg, J., Martin, M., and Donnelly, D.J.** (2014). History and Origin of Russet Burbank (Netted Gem) a Sport of Burbank. *American Journal of Potato Research* **91**: 594–609.

- Braz, G.T., He, L., Zhao, H., Zhang, T., Semrau, K., Rouillard, J.-M., Torres, G.A., and Jiang, J.** (2018). Comparative Oligo-FISH Mapping: An Efficient and Powerful Methodology To Reveal Karyotypic and Chromosomal Evolution. *Genetics* **208**: 513–523.
- Brettell, R.I.S. and Dennis, E.S.** (1991). Reactivation of a silent Ac following tissue culture is associated with heritable alterations in its methylation pattern. *Molecular and General Genetics MGG* **229**: 365–372.
- Carbonell-Bejerano, P., Royo, C., Torres-Pérez, R., Grimplet, J., Fernandez, L., Franco-Zorrilla, J.M., Lijavetzky, D., Baroja, E., Martínez, J., García-Escudero, E., Ibáñez, J., and Martínez-Zapater, J.M.** (2017). Catastrophic Unbalanced Genome Rearrangements Cause Somatic Loss of Berry Color in Grapevine. *Plant Physiol.* **175**: 786–801.
- Carlson, P.S.** (1974). Mitotic crossing-over in a higher plant. *Genet. Res.* **24**: 109–112.
- Čermák, T., Curtin, S.J., Gil-Humanes, J., Čegan, R., Kono, T.J.Y., Konečná, E., Belanto, J.J., Starker, C.G., Mathre, J.W., Greenstein, R.L., and Voytas, D.F.** (2017). A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. *The Plant Cell* **29**: 1196–1217.
- Creissen, G.P. and Karp, A.** (1985). Karyotypic changes in potato plants regenerated from protoplasts. *Plant Cell Tissue Organ Cult.* **4**: 171–182.
- Decaestecker, W., Buono, R.A., Pfeiffer, M.L., Vangheluwe, N., Jourquin, J., Karimi, M., Van Isterdael, G., Beeckman, T., Nowack, M.K., and Jacobs, T.B.** (2019). CRISPR-TSKO: A Technique for Efficient Mutagenesis in Specific Cell Types, Tissues, or Organs in Arabidopsis. *Plant Cell* **31**: 2868–2887.
- Demirer, G.S., Zhang, H., Goh, N.S., González-Grandío, E., and Landry, M.P.** (2019). Carbon nanotube-mediated DNA delivery without transgene integration in intact plants. *Nat. Protoc.* **14**: 2954–2971.
- Doležel, J., Binarová, P., and Lcretti, S.** (1989). Analysis of nuclear DNA content in plant cells by flow cytometry. *Biol. Plant.* **31**: 113–120.
- Dong, F., Song, J., Naess, S.K., Helgeson, J.P., Gebhardt, C., and Jiang, J.** (2000). Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. *Theor. Appl. Genet.* **101**: 1001–1007.
- Fehér, A.** (2015). Somatic embryogenesis — Stress-induced remodeling of plant cell fate. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1849**: 385–402.
- Fish, N. and Karp, A.** (1986). Improvements in regeneration from protoplasts of potato and studies on chromosome stability. *Züchter Genet. Breed. Res.* **72**: 405–412.
- Florentin, A., Damri, M., and Grafi, G.** (2013). Stress induces plant somatic cells to acquire some features of stem cells accompanied by selective chromatin reorganization. *Dev. Dyn.* **242**: 1121–1133.
- Fossi, M., Amundson, K.R., Kuppu, S., Britt, A.B., and Comai, L.** (2019). Regeneration of *Solanum tuberosum* plants from protoplasts induces widespread genome instability. *Plant Physiol.*
- Foster, T.M. and Aranzana, M.J.** (2018). Attention sports fans! The far-reaching contributions of bud sport mutants to horticulture and plant biology. *Hortic Res* **5**: 44.

- Fridman, C. and Koiffmann, C.P.** (2000). Origin of uniparental disomy 15 in patients with Prader-Willi or Angelman syndrome. *Am. J. Med. Genet.* **94**: 249–253.
- Galbraith, D.W., Janda, J., and Lambert, G.M.** (2011). Multiparametric analysis, sorting, and transcriptional profiling of plant protoplasts and nuclei according to cell type. *Methods Mol. Biol.* **699**: 407–429.
- Garrison, E. and Marth, G.** (2012). Haplotype-based variant detection from short-read sequencing. arXiv [q-bio.GN].
- Gernand, D., Golczyk, H., Rutten, T., Ilnicki, T., Houben, A., and Joachimiak, A.J.** (2007). Tissue culture triggers chromosome alterations, amplification, and transposition of repeat sequences in *Allium fistulosum*. *Genome* **50**: 435–442.
- Ghislain M., Zhang D. P., Herrera, M. R. ed** (1999). *Molecular Biology Laboratory Protocols Plant Genotyping: Training Manual* (International Potato Center: Lima, Peru).
- Gill, B.S., Kam-Morgan, L.N.W., and Shepard, J.F.** (1987). Cytogenetic and phenotypic variation in mesophyll cell-derived tetraploid potatoes. *J. Hered.* **78**: 15–20.
- Grafi, G. and Barak, S.** (2015). Stress induces cell dedifferentiation in plants. *Biochim. Biophys. Acta* **1849**: 378–384.
- Guan, J., Xu, Y., Yu, Y., Fu, J., Ren, F., Guo, J., Zhao, J., Jiang, Q., Wei, J., and Xie, H.** (2021). Genome structure variation analyses of peach reveal population dynamics and a 1.67 Mb causal inversion for fruit shape. *Genome Biol.* **22**: 13.
- Guo, W., Comai, L., and Henry, I.M.** (2021). Chromoanagenesis from radiation-induced genome damage in *Populus*. *PLoS Genet.* **17**: e1009735.
- Han, Z., Crisp, P.A., Stelpflug, S., Kaeppler, S.M., Li, Q., and Springer, N.M.** (2018). Heritable Epigenomic Changes to the Maize Methylome Resulting from Tissue Culture. *Genetics* **209**: 983–995.
- Hardigan, M.A. et al.** (2016). Genome Reduction Uncovers a Large Dispensable Genome and Adaptive Role for Copy Number Variation in Asexually Propagated *Solanum tuberosum*. *Plant Cell* **28**: 388–405.
- Hardigan, M.A., Laimbeer, F.P.E., Newton, L., Crisovan, E., Hamilton, J.P., Vaillancourt, B., Wiegert-Rininger, K., Wood, J.C., Douches, D.S., Farré, E.M., Veilleux, R.E., and Buell, C.R.** (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. U. S. A.* **114**: E9999–E10008.
- He, L., Braz, G.T., Torres, G.A., and Jiang, J.** (2018). Chromosome painting in meiosis reveals pairing of specific chromosomes in polyploid *Solanum* species. *Chromosoma* **127**: 505–513.
- Henry, I.M., Zinkgraf, M.S., Groover, A.T., and Comai, L.** (2015). A System for Dosage-Based Functional Genomics in Poplar. *Plant Cell* **27**: 2370–2383.
- Hermesen, J.G.T. and Verdenius, J.** (1973). Selection from *Solanum tuberosum* group Phureja of genotypes combining high-frequency haploid induction with homozygosity for embryo-spot. *Euphytica* **22**: 244–259.

- Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H., and Kanda, M.** (1996). Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 7783–7788.
- Hirsch, C.N., Hirsch, C.D., Felcher, K., Coombs, J., Zarka, D., Van Deynze, A., De Jong, W., Veilleux, R.E., Jansky, S., Bethke, P., Douches, D.S., and Buell, C.R.** (2013). Retrospective view of North American potato (*Solanum tuberosum* L.) breeding in the 20th and 21st centuries. *G3* **3**: 1003–1013.
- Holland, A.J. and Cleveland, D.W.** (2012). Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements. *Nat. Med.* **18**: 1630–1638.
- Howard, H.W.** (1972). The stability of an L3 mutant potato chimera. *Potato Research* **15**: 374–377.
- Howard, H.W.** (1971). The stability of L1-mutant periclinal potato chimeras. *Potato Res.* **14**: 91–93.
- Howard, H.W., Wainwright, J., and Fuller, J.M.** (1964). The number of independent layers at the stem apex in potatoes. *Genetica* **34**: 113–120.
- Ikeuchi, M., Sugimoto, K., and Iwase, A.** (2013). Plant callus: mechanisms of induction and repression. *Plant Cell* **25**: 3159–3173.
- Jiang, C., Mithani, A., Gan, X., Belfield, E.J., Klingler, J.P., Zhu, J.-K., Ragoussis, J., Mott, R., and Harberd, N.P.** (2011). Regenerant Arabidopsis lineages display a distinct genome-wide spectrum of mutations conferring variant phenotypes. *Curr. Biol.* **21**: 1385–1390.
- Johnsson, M., Gaynor, R.C., Jenko, J., Gorjanc, G., de Koning, D.-J., and Hickey, J.M.** (2019). Removal of alleles by genome editing (RAGE) against deleterious load. *Genet. Sel. Evol.* **51**: 14.
- Jong, H.D. and De Jong, H.** (1981). Inheritance of russeting in cultivated diploid potatoes. *Potato Research* **24**: 309–313.
- Jung, C.S., Griffiths, H.M., De Jong, D.M., Cheng, S., Bodis, M., and De Jong, W.S.** (2005). The potato P locus codes for flavonoid 3', 5'-hydroxylase. *Theor. Appl. Genet.* **110**: 269–275.
- Jung, C.S., Griffiths, H.M., De Jong, D.M., Cheng, S., Bodis, M., Kim, T.S., and De Jong, W.S.** (2009). The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. *Theor. Appl. Genet.* **120**: 45–57.
- Kaeppeler, S.M. and Phillips, R.L.** (1993a). DNA methylation and tissue culture-induced variation in plants. *In Vitro Cellular & Developmental Biology - Plant* **29**: 125–130.
- Kaeppeler, S.M. and Phillips, R.L.** (1993b). Tissue culture-induced DNA methylation variation in maize. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 8773–8776.
- Karp, A., Jones, M.G.K., Ooms, G., and Bright, S.W.J.** (1987). Potato Protoplasts and Tissue Culture in Crop Improvement. *Biotechnol. Genet. Eng. Rev.* **5**: 1–32.
- Karp, A., Nelson, R.S., Thomas, E., and Bright, S.W.** (1982). Chromosome variation in protoplast-derived potato plants. *Theor. Appl. Genet.* **63**: 265–272.
- Kato, A., Lamb, J.C., and Birchler, J.A.** (2004). Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 13554–13559.

- Kelliher, T. et al.** (2019). One-step genome editing of elite crop germplasm during haploid induction. *Nat. Biotechnol.* **37**: 287–292.
- Kikuchi, K., Terauchi, K., Wada, M., and Hirano, H.-Y.** (2003). The plant MITE mPing is mobilized in anther culture. *Nature* **421**: 167–170.
- Kim, M.-S. et al.** (2021). The patterns of deleterious mutations during the domestication of soybean. *Nat. Commun.* **12**: 97.
- Korbel, J.O. and Campbell, P.J.** (2013). Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**: 1226–1236.
- Laimbeer, F.P.E., Bargmann, B.O.R., Holt, S.H., Pratt, T., Peterson, B., Doulis, A.G., Buell, C.R., and Veilleux, R.E.** (2020). Characterization of the F Locus Responsible for Floral Anthocyanin Production in Potato. *G3* **10**: 3871–3879.
- Larkin, P.J., Ryan, S.A., Brettell, R.I., and Scowcroft, W.R.** (1984). Heritable somaclonal variation in wheat. *Theor. Appl. Genet.* **67**: 443–455.
- Larkin, P.J. and Scowcroft, W.R.** (1981). Somaclonal variation - a novel source of variability from cell cultures for plant improvement. *Züchter Genet. Breed. Res.* **60**: 197–214.
- Lee, M. and Phillips, R.L.** (1988). The chromosomal basis of somaclonal variation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **39**: 413–437.
- Leever, G., Trank, W.A., Shaver, G., Miller, J.C., and Pavlista, A.D.** (1994). Norgold russet, superior and red la soda strains selected for potato cultivar improvement in Nebraska. *Am. Potato J.* **71**: 133–143.
- Legay, S., Guerriero, G., André, C., Guignard, C., Cocco, E., Charton, S., Boutry, M., Rowland, O., and Hausman, J.-F.** (2016). MdMyb93 is a regulator of suberin deposition in russeted apple fruit skins. *New Phytol.* **212**: 977–991.
- van Lieshout, N., van der Burgt, A., de Vries, M.E., Ter Maat, M., Eickholt, D., Esselink, D., van Kaauwen, M.P.W., Kodde, L.P., Visser, R.G.F., Lindhout, P., and Finkers, R.** (2020). Solyntus, the New Highly Contiguous Reference Genome for Potato (*Solanum tuberosum*). *G3* **10**: 3489–3495.
- Li, H., Li, J., Chen, J., Yan, L., and Xia, L.** (2020). Precise Modifications of Both Exogenous and Endogenous Genes in Rice by Prime Editing. *Mol. Plant* **13**: 671–674.
- Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., Wang, Y., Anzalone, A.V., Raguram, A., Doman, J.L., Liu, D.R., and Gao, C.** (2020). Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**: 582–585.
- Lozano, R. et al.** (2021). Comparative evolutionary genetics of deleterious load in sorghum and maize. *Nat Plants* **7**: 17–24.
- Maher, M.F., Nasti, R.A., Vollbrecht, M., Starker, C.G., Clark, M.D., and Voytas, D.F.** (2019). Plant gene editing through de novo induction of meristems. *Nat. Biotechnol.* **38**: 84–89.
- Martin, M.** (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12.

- Mascari, M.J., Gottlieb, W., Rogan, P.K., Butler, M.G., Waller, D.A., Armour, J.A., Jeffreys, A.J., Ladda, R.L., and Nicholls, R.D.** (1992). The frequency of uniparental disomy in Prader-Willi syndrome. Implications for molecular diagnosis. *N. Engl. J. Med.* **326**: 1599–1607.
- Matern, U., Strobel, G., and Shepard, J.** (1978). Reaction to phytotoxins in a potato population derived from mesophyll protoplasts. *Proc. Natl. Acad. Sci. U. S. A.* **75**: 4935–4939.
- Miller, J.C.** (1954). Selection of desirable somatic mutations; A means of potato improvement. *Am. Potato J.* **31**: 358–359.
- Miller, J.C., Scheuring, D.C., Miller, J.P., and Fernandez, G.C.J.** (1999). Selection, evaluation, and identification of improved Russet Norkotah strains. *Am. J. Potato Res.* **76**: 161–167.
- Miyao, A., Nakagome, M., Ohnuma, T., Yamagata, H., Kanamori, H., Katayose, Y., Takahashi, A., Matsumoto, T., and Hirochika, H.** (2012). Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. *Plant Cell Physiol.* **53**: 256–264.
- Nassar, A.M.K., Ortiz-Medina, E., Leclerc, Y., and Donnelly, D.J.** (2008). Periclinal chimeral status of New Brunswick “russet Burbank” potato. *Am. J. Potato Res.* **85**: 432–437.
- Nasti, R.A. and Voytas, D.F.** (2021). Attaining the promise of plant gene editing at scale. *Proc. Natl. Acad. Sci. U. S. A.* **118**.
- Neelakandan, A.K. and Wang, K.** (2012). Recent progress in the understanding of tissue culture-induced genome level changes in plants and potential applications. *Plant Cell Rep.* **31**: 597–620.
- Ordoñez, B., Santayana, M., Aponte, M., Henry, I.M., Comai, L., Eyzaguirre, R., Lindqvist-Kreuzer, H., and Bonierbale, M.** (2021). PL-4 (CIP596131.4): an Improved Potato Haploid Inducer. *Am. J. Potato Res.* **98**: 255–262.
- Pâques, F. and Haber, J.E.** (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**: 349–404.
- Park, J.-S., Park, J.-H., Kim, S.-J., and Park, Y.-D.** (2020). Genome analysis of tissue culture-derived variations in regenerated *Brassica rapa* ssp. *pekinensis* plants using next-generation sequencing. *Horticulture, Environment, and Biotechnology* **61**: 549–558.
- Pellestor, F.** (2019). Chromoanagenesis: cataclysms behind complex chromosomal rearrangements. *Mol. Cytogenet.* **12**: 6.
- Pellestor, F. and Gatinois, V.** (2018). Chromoanasythesis: another way for the formation of complex chromosomal abnormalities in human reproduction. *Hum. Reprod.* **33**: 1381–1387.
- Pelsy, F., Dumas, V., Bévillacqua, L., Hocquigny, S., and Merdinoglu, D.** (2015). Chromosome replacement and deletion lead to clonal polymorphism of berry color in grapevine. *PLoS Genet.* **11**: e1005081.
- Peschke, V.M. and Phillips, R.L.** (1991). Activation of the maize transposable element Suppressor-mutator (Spm) in tissue culture. *Theor. Appl. Genet.* **81**: 90–97.
- Peschke, V.M., Phillips, R.L., and Gengenbach, B.G.** (1987). Discovery of Transposable Element Activity Among Progeny of Tissue Culture—Derived Maize Plants. *Science*.

- Pham, G.M., Hamilton, J.P., Wood, J.C., Burke, J.T., Zhao, H., Vaillancourt, B., Ou, S., Jiang, J., and Buell, C.R.** (2020). Construction of a chromosome-scale long-read reference genome assembly for potato. *Gigascience* **9**.
- Pham, G.M., Newton, L., Wiegert-Rininger, K., Vaillancourt, B., Douches, D.S., and Buell, C.R.** (2017). Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. *Plant J.* **92**: 624–637.
- Phillips, R.L., Kaeppeler, S.M., and Olhoft, P.** (1994). Genetic instability of plant tissue cultures: breakdown of normal controls. *Proc. Natl. Acad. Sci. U. S. A.* **91**: 5222–5226.
- Prodhomme, C., Vos, P.G., Paulo, M.J., Tammes, J.E., Visser, R.G.F., Vossen, J.H., and van Eck, H.J.** (2020). Distribution of P1(D1) wart disease resistance in potato germplasm and GWAS identification of haplotype-specific SNP markers. *Theor. Appl. Genet.* **133**: 1859–1871.
- Pucker, B., Rückert, C., Stracke, R., Viehöver, P., Kalinowski, J., and Weisshaar, B.** (2019). Twenty-Five Years of Propagation in Suspension Cell Culture Results in Substantial Alterations of the Arabidopsis Thaliana Genome. *Genes* **10**.
- Ramstein, G.P. and Buckler, E.S.** (2021). Prediction of evolutionary constraint by genomic annotations improves prioritization of causal variants in maize. *bioRxiv*: 2021.09.03.458856.
- Ramulu, K.S., Sree Ramulu, K., Dijkhuis, P., and Roest, S.** (1983). Phenotypic variation and ploidy level of plants regenerated from protoplasts of tetraploid potato (*Solanum tuberosum* L. cv. “Bintje”). *Theoretical and Applied Genetics* **65**: 329–338.
- Ramu, P., Esuma, W., Kawuki, R., Rabbi, I.Y., Egesi, C., Bredeson, J.V., Bart, R.S., Verma, J., Buckler, E.S., and Lu, F.** (2017). Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nat. Genet.* **49**: 959–963.
- Robinson, W.P.** (2000). Mechanisms leading to uniparental disomy and their clinical consequences. *Bioessays* **22**: 452–459.
- Sato, M., Hosokawa, M., and Doi, M.** (2011a). Somaclonal Variation Is Induced De Novo via the Tissue Culture Process: A Study Quantifying Mutated Cells in Saintpaulia. *PLoS ONE* **6**: e23541.
- Sato, M., Kawabe, T., Hosokawa, M., Tatsuzawa, F., and Doi, M.** (2011b). Tissue culture-induced flower-color changes in Saintpaulia caused by excision of the transposon inserted in the flavonoid 3', 5' hydroxylase (F3'5'H) promoter. *Plant Cell Rep.* **30**: 929–939.
- Schmeing, S. and Robinson, M.D.** (2021). ReSeq simulates realistic Illumina high-throughput sequencing data. *Genome Biol.* **22**: 67.
- Schubert, I., Schubert, V., and Fuchs, J.** (2011). No evidence for “break-induced replication” in a higher plant—but break-induced conversion may occur. *Front. Plant Sci.* **2**: 8.
- Sebastiani, L., Lenzi, A., Pugliesi, C., and Fambrini, M.** (1994). Somaclonal variation for resistance to *Verticillium dahliae* in potato (*Solanum tuberosum* L.) plants regenerated from callus. *Euphytica* **80**: 5–11.
- Secor, G.A. and Shepard, J.F.** (1981). Variability of Protoplast-derived Potato Clones I. *Crop Science* **21**: 102.

- Sharma, S.K., MacKenzie, K., McLean, K., Dale, F., Daniels, S., and Bryan, G.J.** (2018). Linkage Disequilibrium and Evaluation of Genome-Wide Association Mapping Models in Tetraploid Potato. *G3* **8**: 3185–3202.
- Shepard, J.F., Bidney, D., and Shahin, E.** (1980). Potato protoplasts in crop improvement. *Science* **208**: 17–24.
- Sree Ramulu, K., Dijkhuis, P., Roest, S., Bokelmann, G.S., and De Groot, B.** (1984). Early occurrence of genetic instability in protoplast cultures of potato. *Plant Sci. Lett.* **36**: 79–86.
- Sref Ramulu, K., Dijkhuis, P., Roest, S., Bokelmann, G.S., and Groot, B.** (1986). Variation in phenotype and chromosome number of plants regenerated from protoplasts of dihaploid and tetraploid potato. *Plant Breed.* **97**: 119–128.
- Stelpflug, S.C., Eichten, S.R., Hermanson, P.J., Springer, N.M., and Kaeppler, S.M.** (2014). Consistent and heritable alterations of DNA methylation are induced by tissue culture in maize. *Genetics* **198**: 209–218.
- Stephens, P.J. et al.** (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**: 27–40.
- Stroud, H., Ding, B., Simon, S.A., Feng, S., Bellizzi, M., Pellegrini, M., Wang, G.-L., Meyers, B.C., and Jacobsen, S.E.** (2013). Plants regenerated from tissue culture contain stable epigenome changes in rice. *Elife* **2**: e00354.
- Sun, H., Jiao, W.B., Krause, K., Campoy, J.A., and Goel, M.** (2021). Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *bioRxiv*.
- Tan, E.H., Henry, I.M., Ravi, M., Bradnam, K.R., Mandakova, T., Marimuthu, M.P., Korf, I., Lysak, M.A., Comai, L., and Chan, S.W.** (2015). Catastrophic chromosomal restructuring during genome elimination in plants. *Elife* **4**.
- Tang, X. et al.** (2018). A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biol.* **19**: 84.
- Tang, X. et al.** (2020). Plant Prime Editors Enable Precise Gene Editing in Rice Cells. *Mol. Plant* **13**: 667–670.
- Tan, Q., Liu, X., Gao, H., Xiao, W., Chen, X., Fu, X., Li, L., Li, D., and Gao, D.** (2019). Comparison Between Flat and Round Peaches, Genomic Evidences of Heterozygosity Events. *Front. Plant Sci.* **10**: 592.
- Tegg, R.S., Thangavel, T., Aminian, H., and Wilson, C.R.** (2013). Somaclonal selection in potato for resistance to common scab provides concurrent resistance to powdery scab. *Plant Pathol.* **62**: 922–931.
- Thieme, R. and Griess, H.** (2005). Somaclonal variation in tuber traits of potato. *Potato Res.* **48**: 153–165.
- Thomas, E., Bright, S.W., Franklin, J., Lancaster, V.A., Mifflin, B.J., and Gibson, R.** (1982). Variation amongst protoplast-derived potato plants (*Solanum tuberosum* cv. “Maris Bard”). *Theor. Appl. Genet.* **62**: 65–68.

- Umbreit, N.T., Zhang, C.-Z., Lynch, L.D., Blaine, L.J., Cheng, A.M., Tourdot, R., Sun, L., Alnubarak, H.F., Judge, K., Mitchell, T.J., Spektor, A., and Pellman, D.** (2020). Mechanisms generating cancer genome complexity from a single cell division error. *Science* **368**.
- Veilleux, R.E. and Johnson, A.A.T.** (2010). Somaclonal variation: Molecular analysis, transformation interaction, and utilization. In *Plant Breeding Reviews* (John Wiley & Sons, Inc.: Oxford, UK), pp. 229–268.
- Vining, K., Pomraning, K.R., Wilhelm, L.J., Ma, C., Pellegrini, M., Di, Y., Mockler, T.C., Freitag, M., and Strauss, S.H.** (2013). Methylome reorganization during in vitro dedifferentiation and regeneration of *Populus trichocarpa*. *BMC Plant Biol.* **13**: 92.
- Wang, L. et al.** (2021). Somatic variations led to the selection of acidic and acidless orange cultivars. *Nat Plants* **7**: 954–965.
- Wheeler, V.A., Evans, N.E., Foulger, D., Webb, K.J., Karp, A., Franklin, J., and Bright, S.W.J.** (1985). Shoot Formation from Explant Cultures of Fourteen Potato Cultivars and Studies of the Cytology and Morphology of Regenerated Plants. *Ann. Bot.* **55**: 309–320.
- Wijnker, E. et al.** (2013). The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. *Elife* **2**: e01426.
- Wilson, C.R., Tegg, R.S., and Hingston, L.H.** (2010). Yield and cooking qualities of somaclonal variants of cv. Russet Burbank selected for resistance to common scab disease of potato. *Annals of Applied Biology* **157**: 283–297.
- Wolter, F., Klemm, J., and Puchta, H.** (2018). Efficient in planta gene targeting in *Arabidopsis* using egg cell-specific expression of the Cas9 nuclease of *Staphylococcus aureus*. *Plant J.* **94**: 735–746.
- Yuan, S., Kawasaki, S., Abdellatif, I.M.Y., Nishida, K., Kondo, A., Ariizumi, T., Ezura, H., and Miura, K.** (2021). Efficient base editing in tomato using a highly expressed transient system. *Plant Cell Rep.* **40**: 667–676.
- Zhang, C., Wang, P., Tang, D., Yang, Z., Lu, F., Qi, J., Tawari, N.R., Shang, Y., Li, C., and Huang, S.** (2019). The genetic basis of inbreeding depression in potato. *Nat. Genet.* **51**: 374–378.
- Zhang, D. et al.** (2014). Tissue culture-induced heritable genomic variation in rice, and their phenotypic implications. *PLoS One* **9**: e96879.
- Zhang, Y., Cheng, S., De Jong, D., Griffiths, H., Halitschke, R., and De Jong, W.** (2009). The potato R locus codes for dihydroflavonol 4-reductase. *Theor. Appl. Genet.* **119**: 931–937.
- Zhou, Q., Tang, D., Huang, W., Yang, Z., Zhang, Y., Hamilton, J.P., Visser, R.G.F., Bachem, C.W.B., Robin Buell, C., Zhang, Z., Zhang, C., and Huang, S.** (2020). Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nat. Genet.* **52**: 1018–1023.

Figures

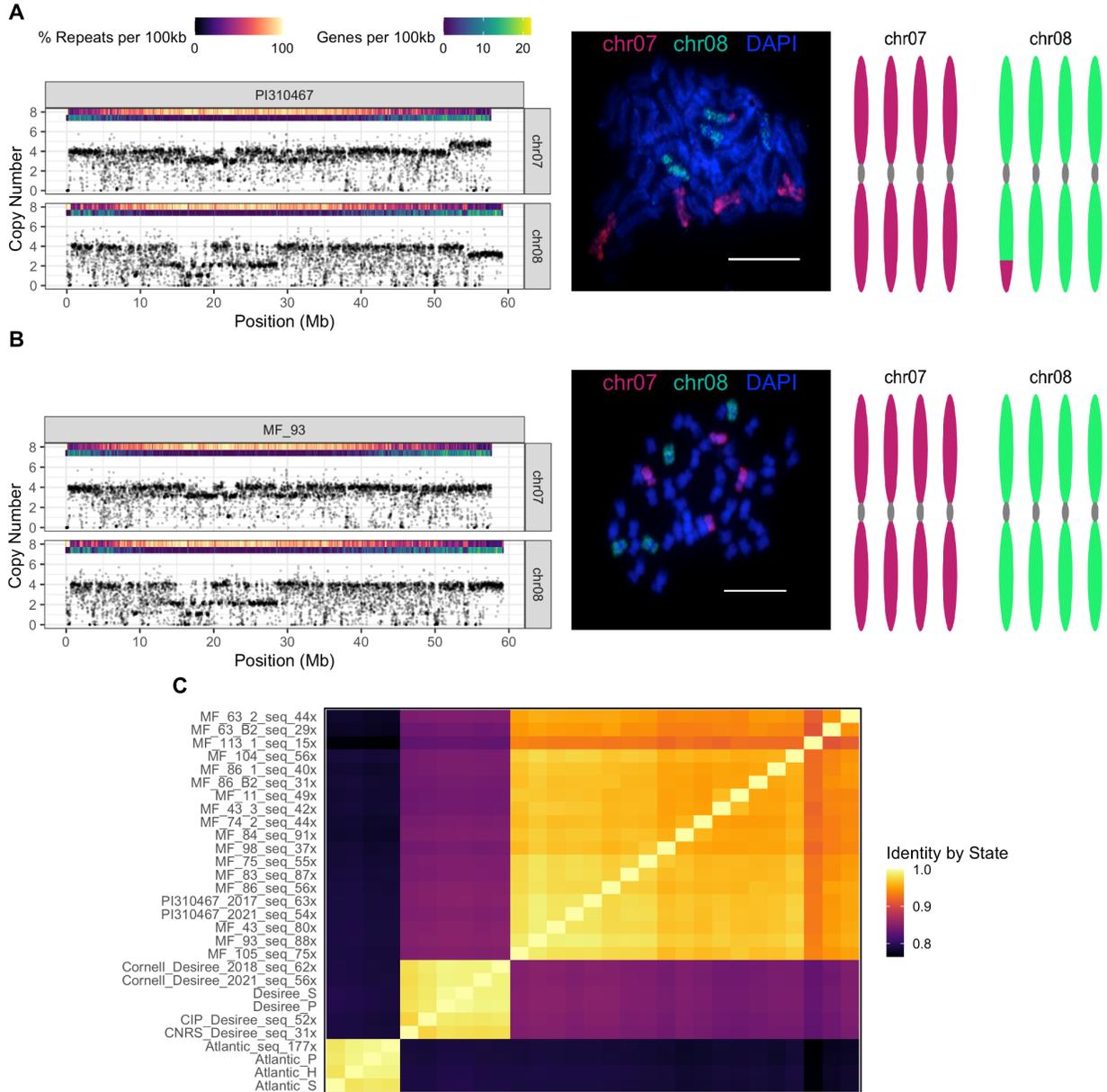


Figure 4.1. Somatic loss of unbalanced translocation tr8-7 among protoplast regenerants of PI310467. A-B) Characterization of translocation presence in PI310467 (A) and absence in regenerant MF93 (B). Left: Coverage plots, with data points corresponding to median read depth in non-overlapping 10kb bins plotted at high transparency. Densities of repeats (% repeat sequence per 100kb, upper stripe) and genes (number of genes per 100kb, lower stripe) according to the DM1-3 v6.1 annotation delineate gene-rich parts of chromosome arms and repeat-rich, structurally-variable pericentromeres. Center: Chromosome-specific oligonucleotide FISH of root tip cells. Chromosomes 7 and 8 are labeled red and green, respectively. Bar: 10 μ m. Right: Chromosome arrangement inferred from sequencing and FISH. C) Heatmap illustrating pairwise relatedness between PI310467, all regenerants of PI310467, the closest relatives (Desiree clones) among a panel of 1,019 potatoes, and Atlantic as an outgroup, based on 1,858

SNPs. For each sequenced sample, average read depth is indicated in the sample name. Each array-genotyped sample is suffixed according to the study it was obtained from: “H” for Hirsch et al (2013), “S” for Sharma et al (2018) or “P” for Prodhomme et al (2020).

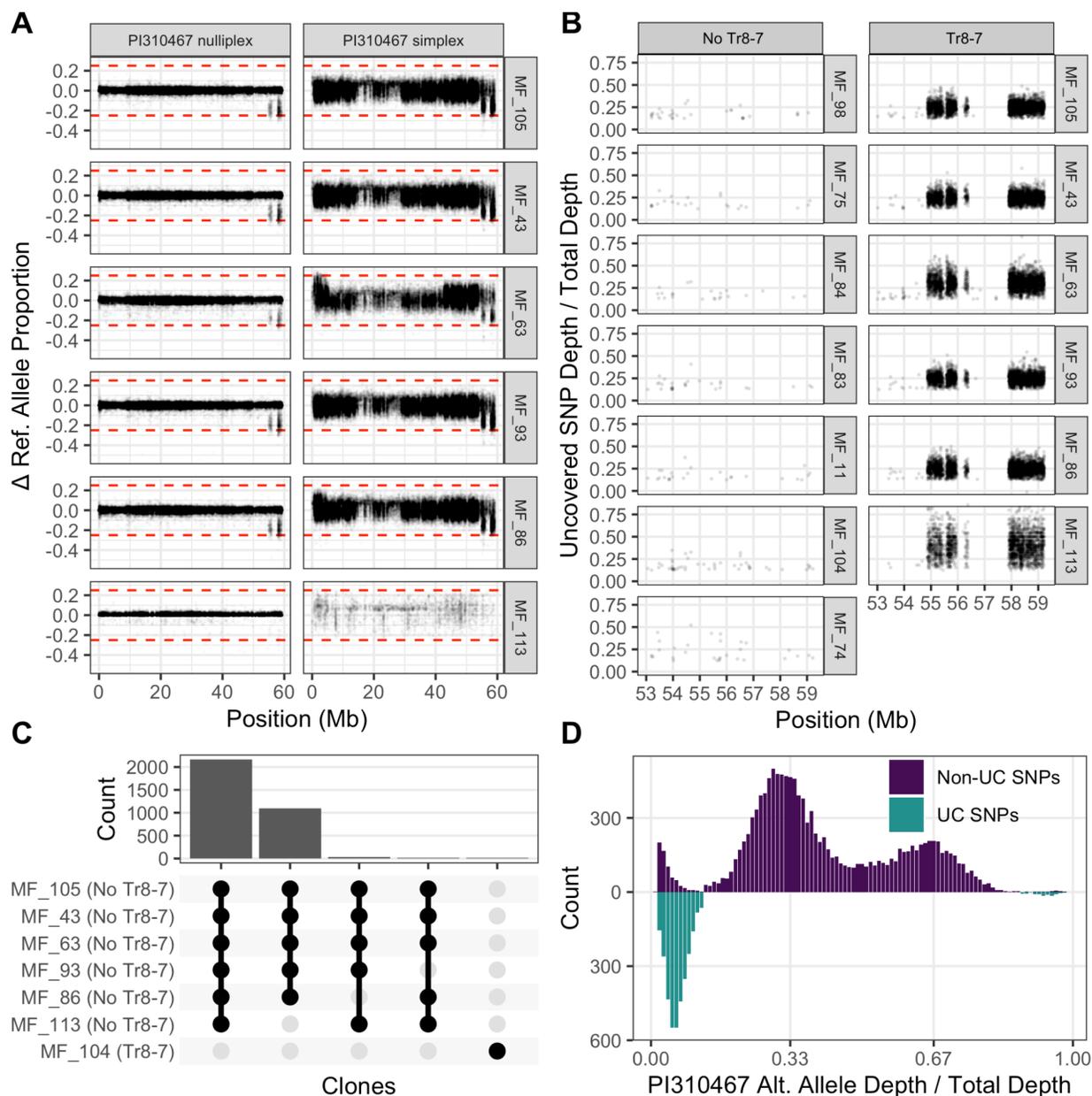


Figure 4.2. Recurring tr8-7 loss due to PI310467 mosaicism. “Uncovered SNPs” define a set of SNPs whose frequency varies between tr8-7-positive and tr8-7-negative regenerants. **A)** Uncovered SNPs among regenerants lacking tr8-7. Dotplot of reference genome position vs. change in SNP locus reference allele proportion between a regenerating and PI310467. Left panel illustrates a 25% decrease in reference allele proportion at PI310467 nulliplex loci distal to the tr8-7 breakpoint. Right panel illustrates a change in reference allele proportion at PI310467 simplex loci proximal to the tr8-7 only in chr08 aneuploids (see Supplemental Fig. S4.5). **B)** Closeup of tr8-7 distal region. Regenerants that lack tr8-7 show uncovered SNPs that are supported by 25% of reads and clustered in the genome. Regenerants that exhibited tr8-7 do not show these clusters. **C)** Upset plot depicting uncovered SNP sharing between clones. For example, the first column indicates that regenerants MF105, MF43, MF63, MF93, MF86 and MF113, which all lack

tr8-7, exhibit the same uncovered SNPs at approximately 2,100 loci. **D)** Histogram of PI310467 alternate allele depth proportion (bin size 0.01) at SNP loci distal to the tr8-7 breakpoint. Purple bars above the X-axis represent loci without an uncovered SNP in any regenerant (non-UC SNPs). Teal bars below the X-axis represent loci with an uncovered allele in at least one regenerant (UC SNPs). UC SNP peaks centered at 0.05 and 0.95 indicate cell subpopulations in PI310467 leaves.

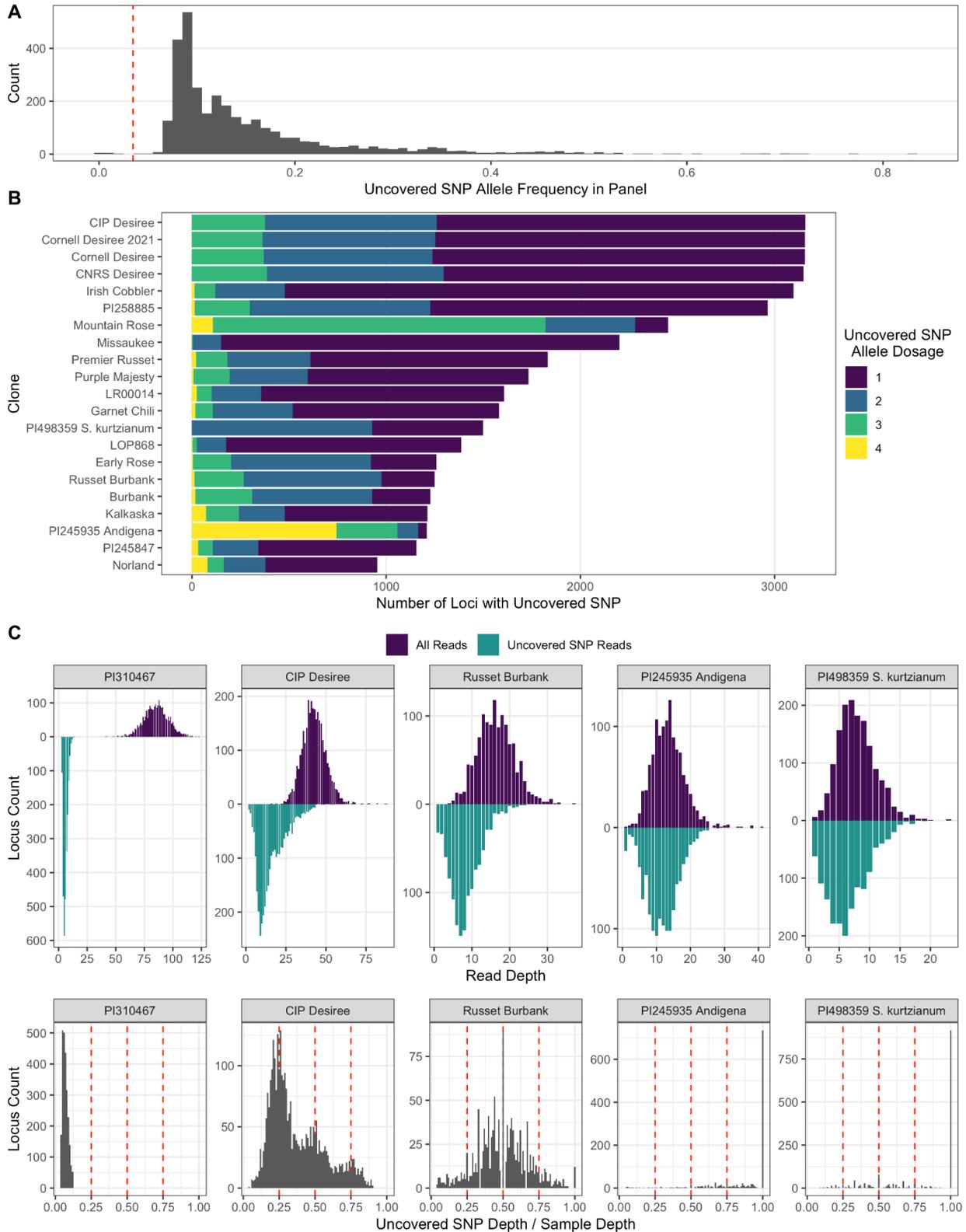


Figure 4.3. SNP alleles associated with somatic loss of tr8-7 loss are common among potatoes. A) Uncovered allele frequency spectrum among a potato diversity panel. Red dashed line: expected allele frequency of a mutation shared by PI310467 regenerants that lack tr8-7 included in the panel. **B)** Bar plot

illustrating the 20 individuals with the greatest number of uncovered SNPs among the panel. Bar color indicates uncovered SNP allele dosage of the clone at hand. C) Uncovered SNP read support of selected clones from panel F. Upper panels: Histograms showing counts of loci by total read depth (bin size 1) in purple, above the Y-axis, and depth of the tr8-7 loss-associated alleles in blue, below the Y-axis, for five selected clones. Lower panels: Histograms of the percentage of uncovered SNPs (bin size 0.01) for five selected clones.

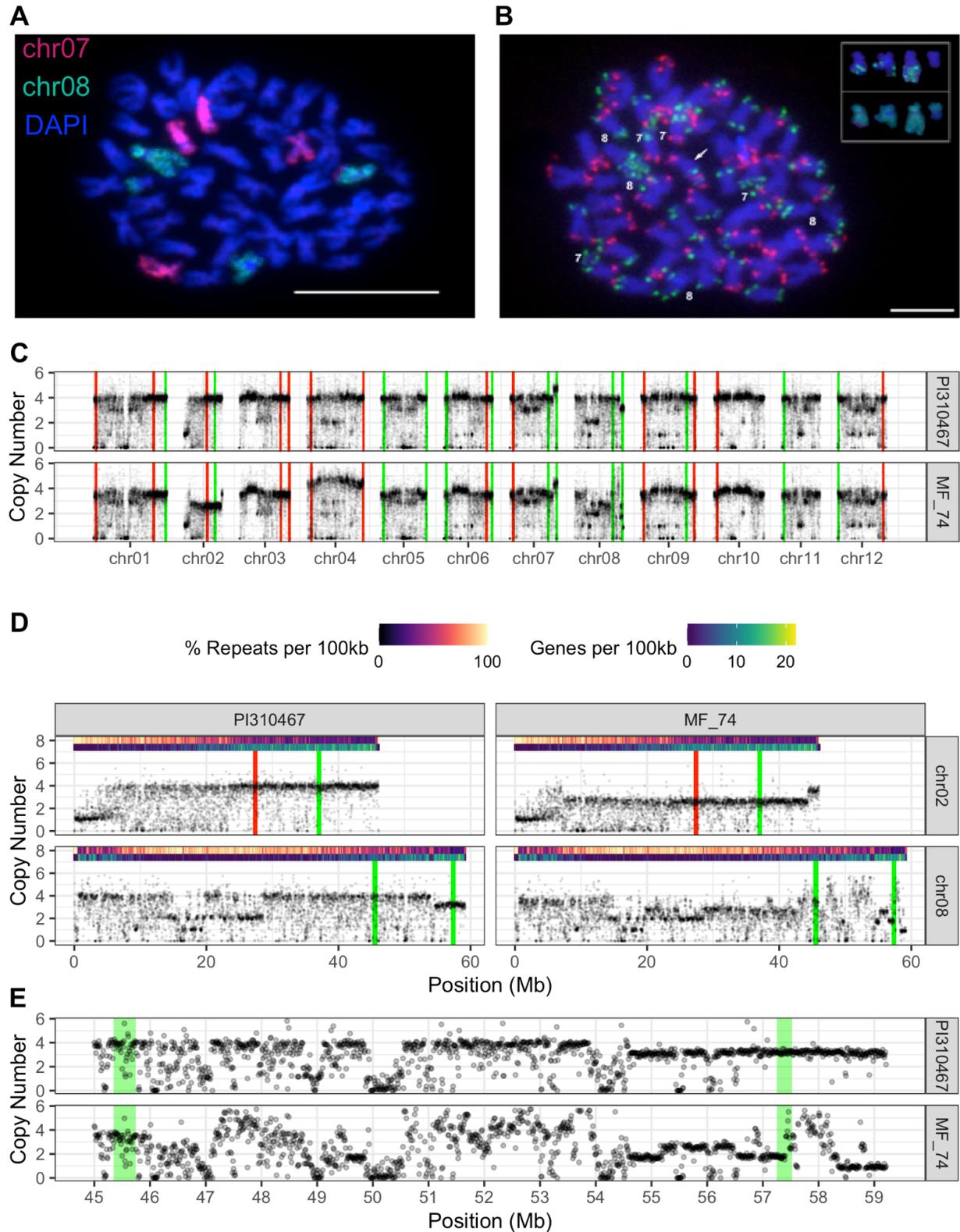


Figure 4.4. Chromosomal rearrangements of regenerant MF74 resemble those seen in chromoanagenesis. A) Chromosome-specific oligonucleotide FISH of MF74 root tip cells. Chromosomes 7 and 8 are labeled red and green, respectively. Bar: 10 μ m. **B)** Barcode FISH mapping of

MF74 root tip cells using the probe sets developed by (Braz et al., 2018). Chromosome 7 and 8 homologs are numbered in the panel. Arrow points to a restructured chromosome. On the top right, chromosome 8 homologs from panels A and B were digitally excised from the panel: barcode FISH mapping is shown on the top row, and chromosome-specific labeling from panel A is shown on the bottom row. Bar: 10 μ m. **C)** Sequencing-derived karyotypes of MF74 and PI310467, showing all 12 chromosomes. Data points correspond to median read depth of non-overlapping 10kb bins and are plotted at high transparency. Green and red regions in both panels indicate DM1-3 v6.1 coordinates of the probes used for labeling panel B, according to color. **D)** Coverage plots for chromosome 2 and 8 for both PI310467 and MF74. Data points correspond to median read depth of non-overlapping 10kb bins and are plotted at high transparency. Densities of repeats (% repeat sequence per 100kb, upper stripe) and genes (number of genes per 100kb, lower stripe) according to the DM1-3 v6.1 annotation are shown above each panel. Green and red regions indicate DM1-3 v6.1 coordinates of the probes used for labeling panel B, according to color. **E)** Closeup of the chromosome 8 region affected by complex structural rearrangement in MF74. Green regions indicate DM1-3 v6.1 coordinates of the probes used for labeling panel B.

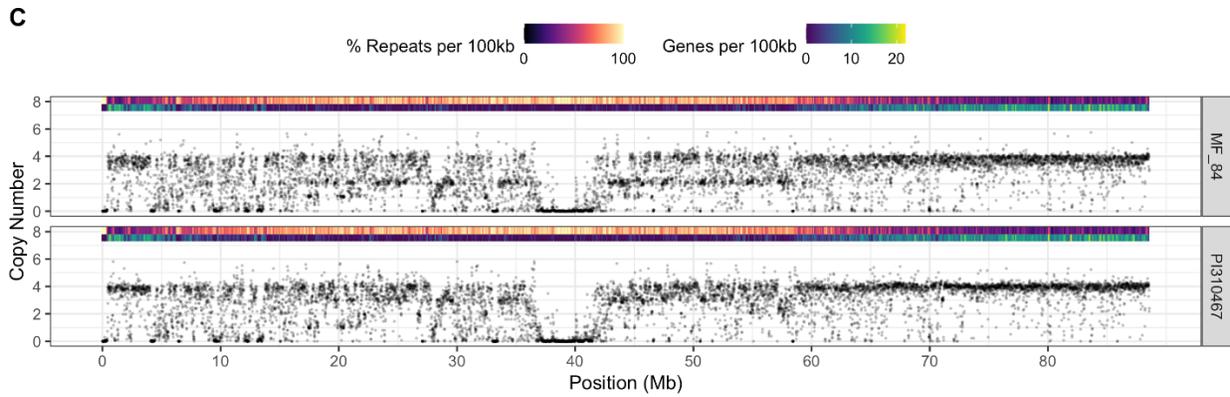
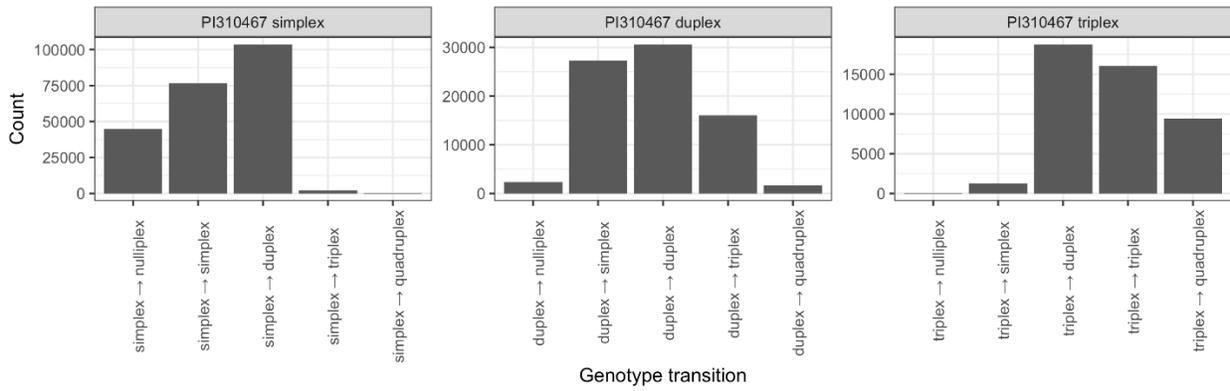
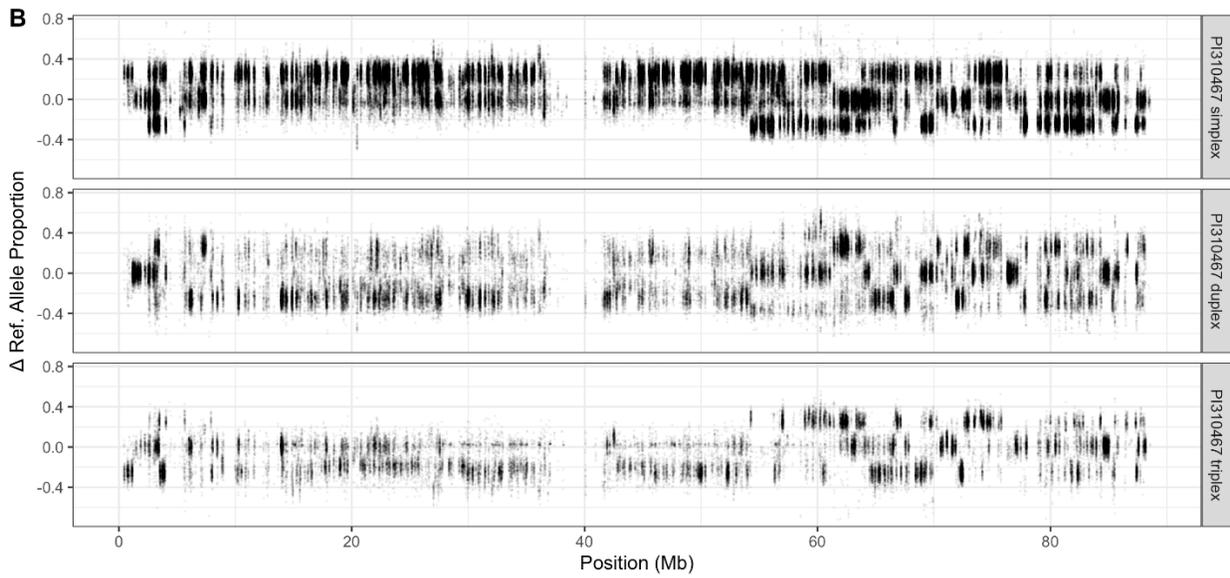
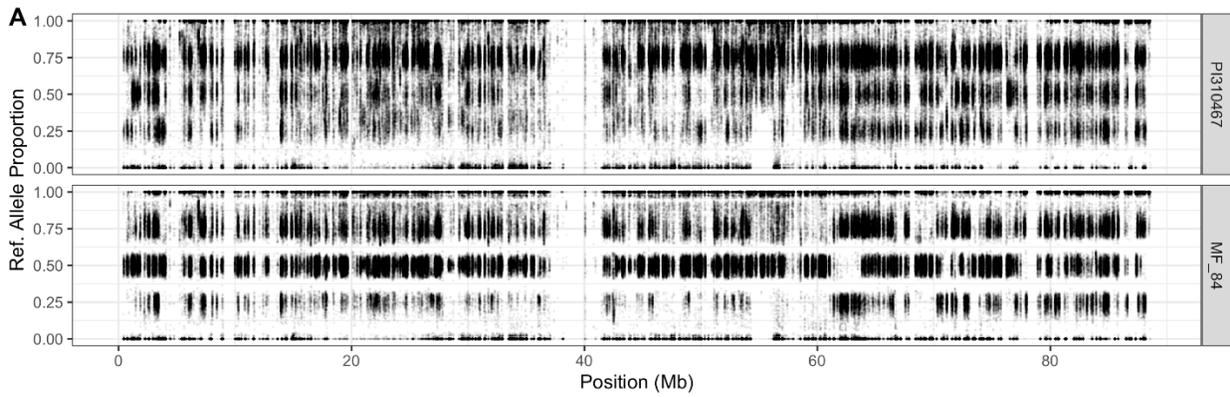


Figure 4.5. Copy-neutral change of heterozygosity consistent with somatic chromosome substitution. For both samples, only chr01 is shown. **A)** Reference allele dosage of PI310467 (upper panel) and regenerant MF84 (lower panel). Only chromosome 1 is shown. Data points correspond to the proportion of reads supporting the reference allele at a given locus and are plotted at high transparency. **B)** Change in reference allele dosage between PI310467 and MF84. Upper panels show dot plots displaying change in allele dosage with respect to reference genome position and organized according to dosage of the non-reference allele (simplex, duplex or triplex). Lower panels show histograms of genotype call concordance between PI310467 and MF84. **C)** Coverage plots of both MF84 and PI310467. Data points correspond to median read depth of non-overlapping 10kb bins and are plotted at high transparency. Above each panel, densities of repeats (% repeat sequence per 100kb, upper stripe) and genes (count of genes per 100kb, lower stripe) based on the DM1-3 v6.1 annotation are shown.

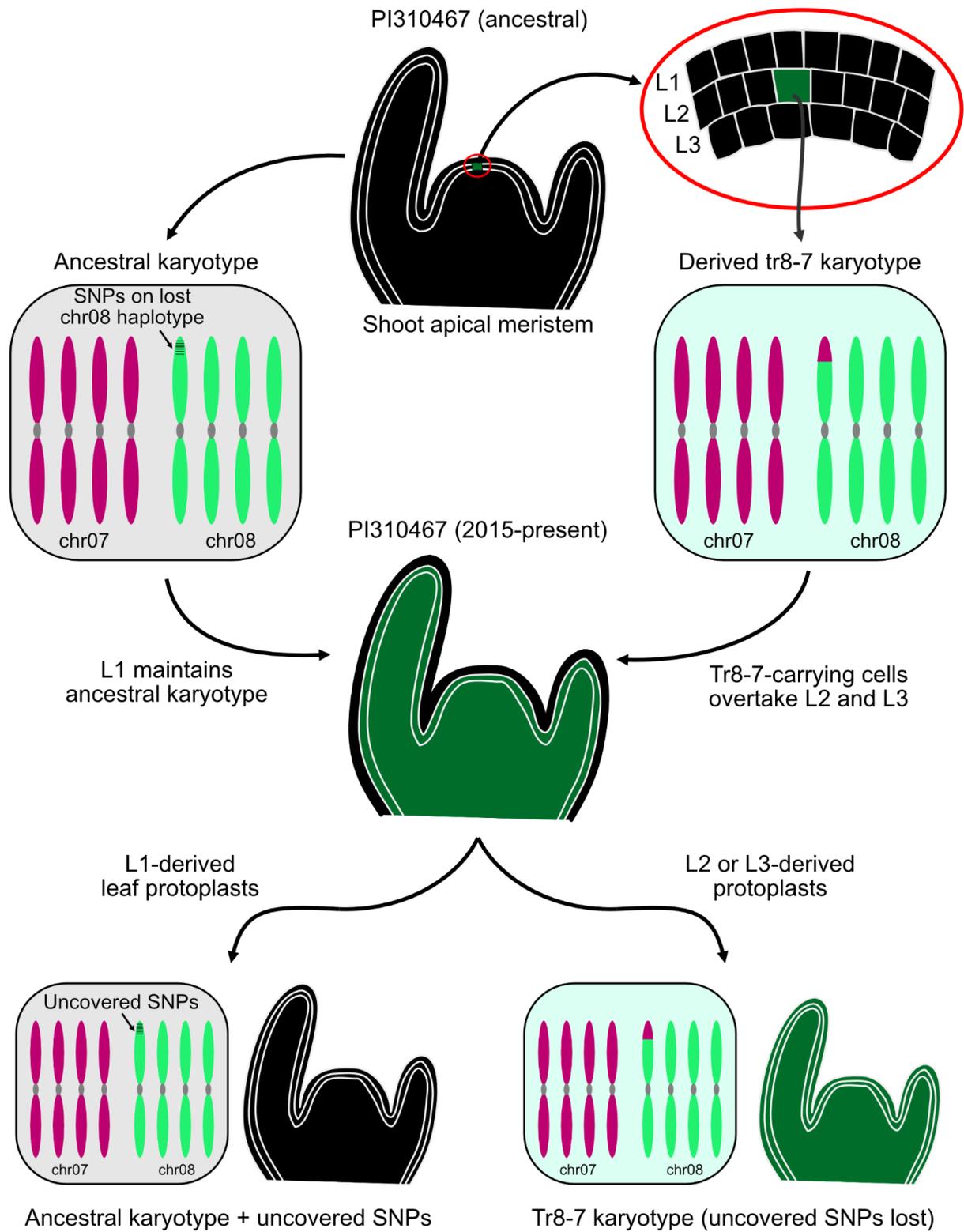
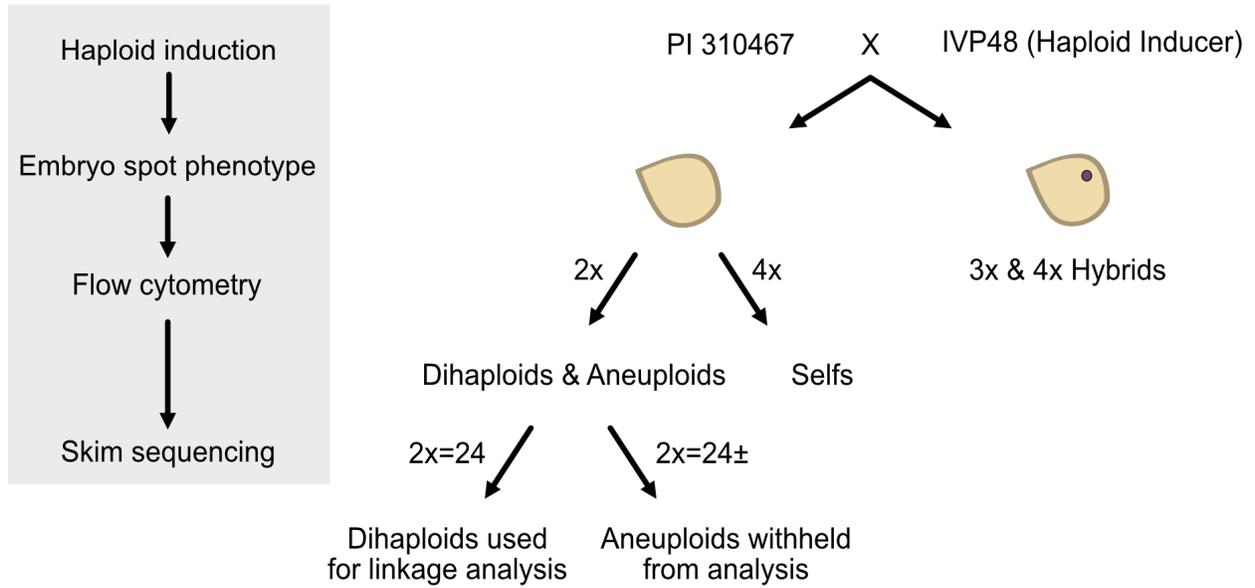


Figure 4.6. Model of chromosomal rearrangement and subsequent developmental events that resulted in PI310467 mosaicism for translocation tr8-7. PI310467 originally carries a balanced karyotype, with one of the four chr08 homologs carrying the uncovered SNPs. An unbalanced

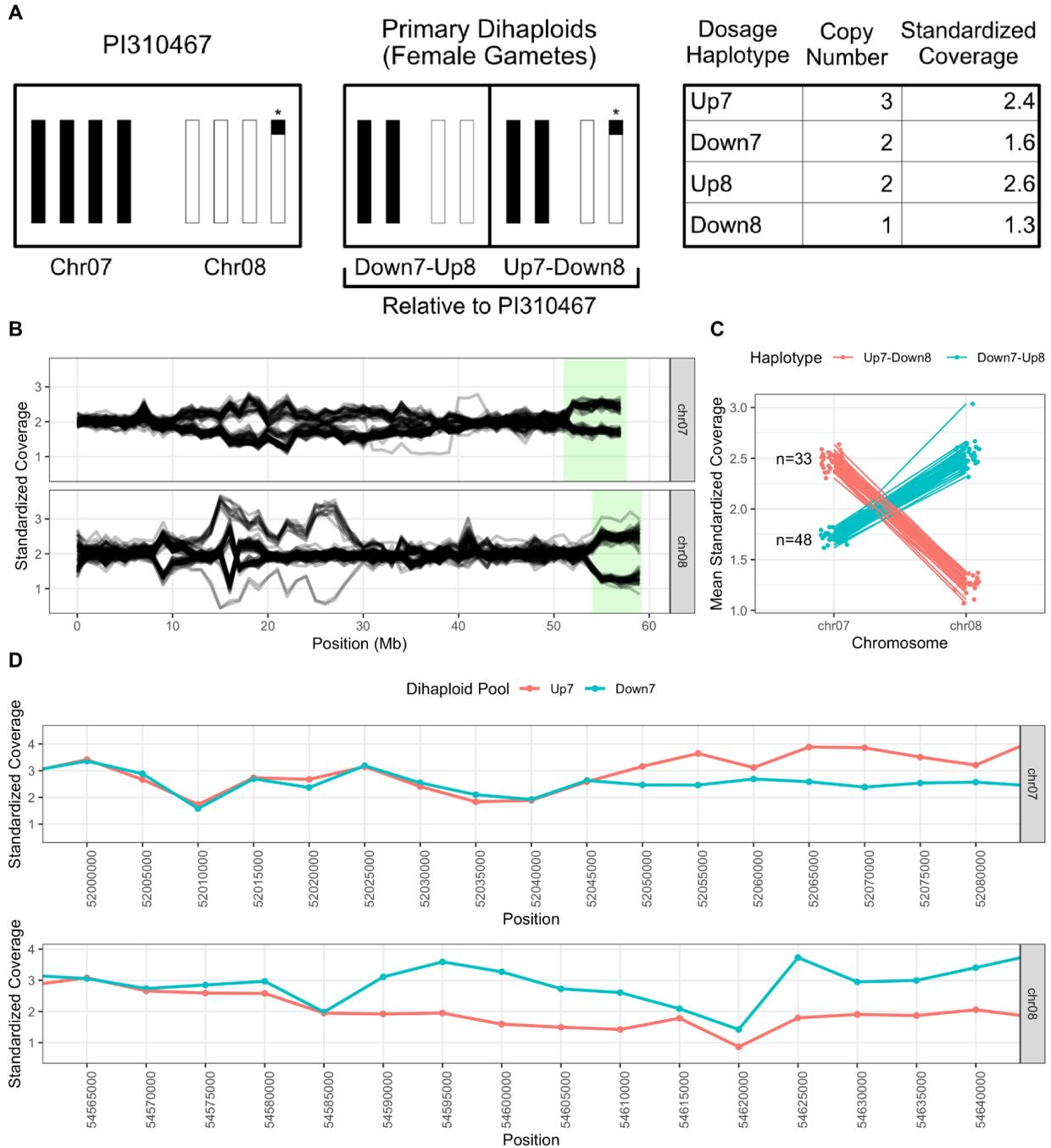
translocation, tr8-7, spontaneously occurs in a single L2-derived cell, resulting in loss of the chr08 haplotype carrying the uncovered SNPs. The cell carrying tr8-7 gives rise to a cell lineage that ultimately becomes fixed in both L2 and L3. In contrast, L1 retains both the ancestral karyotype and the associated chr08 SNPs that were lost from tr8-7-carrying cells. These SNPs will be uncovered by regeneration of L1-derived leaf protoplasts into whole plants. Regeneration of L2- or L3-derived protoplasts results in loss of the ancestral cell lineage.

Supplemental Material

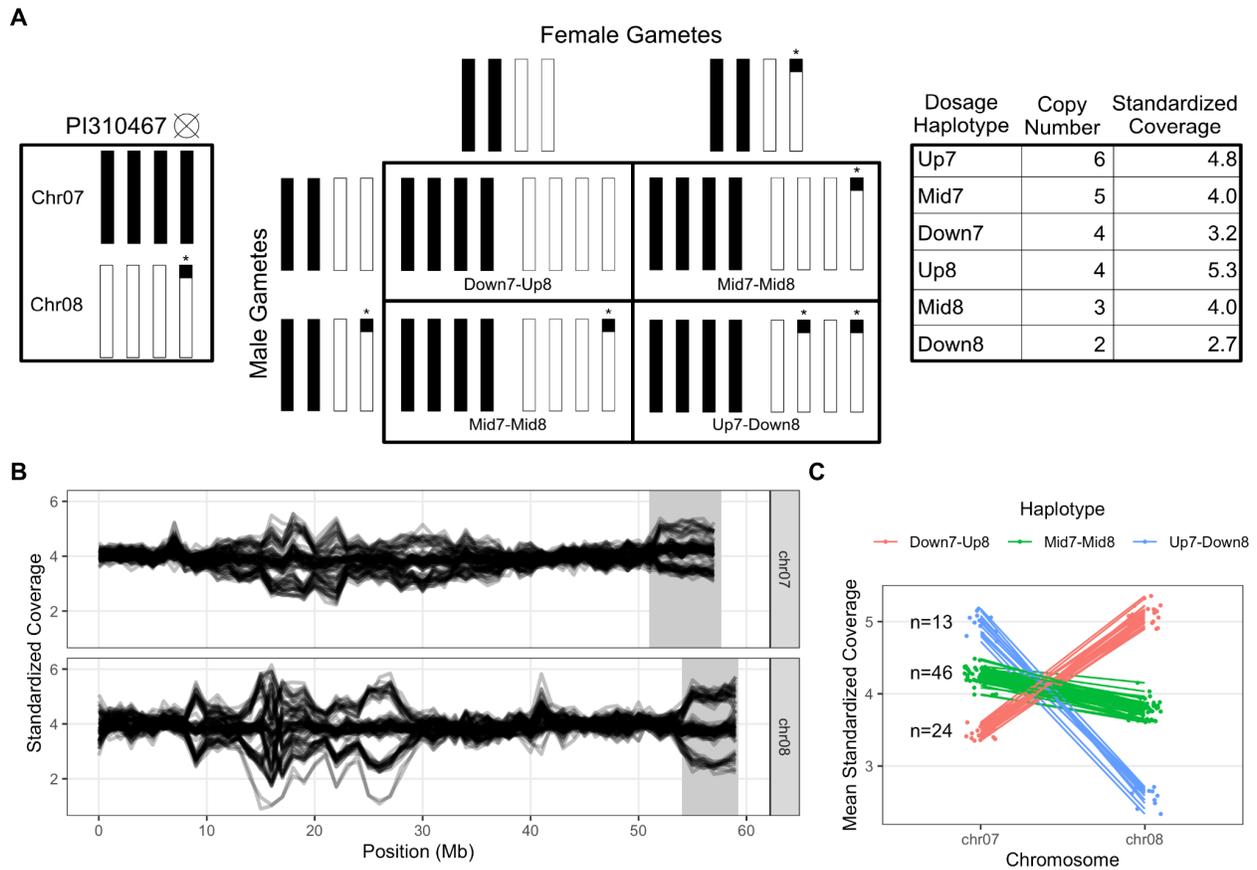
Supplemental Figures



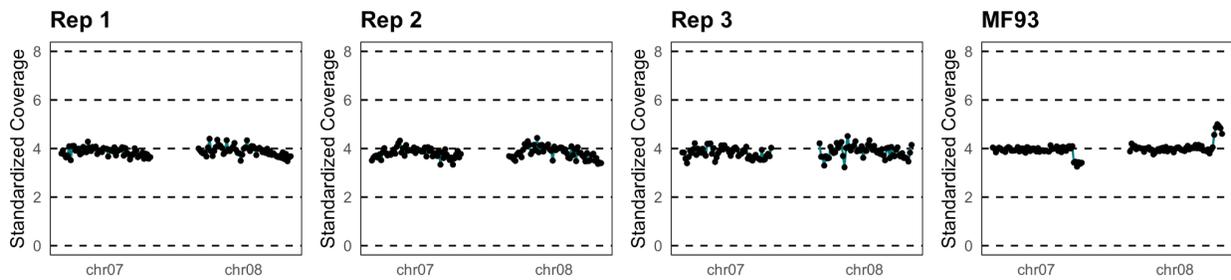
Supplemental Figure S4.1. Workflow for classification of dihaploids, aneuploids, selfs and hybrids. Supplement to Figure 4.1.



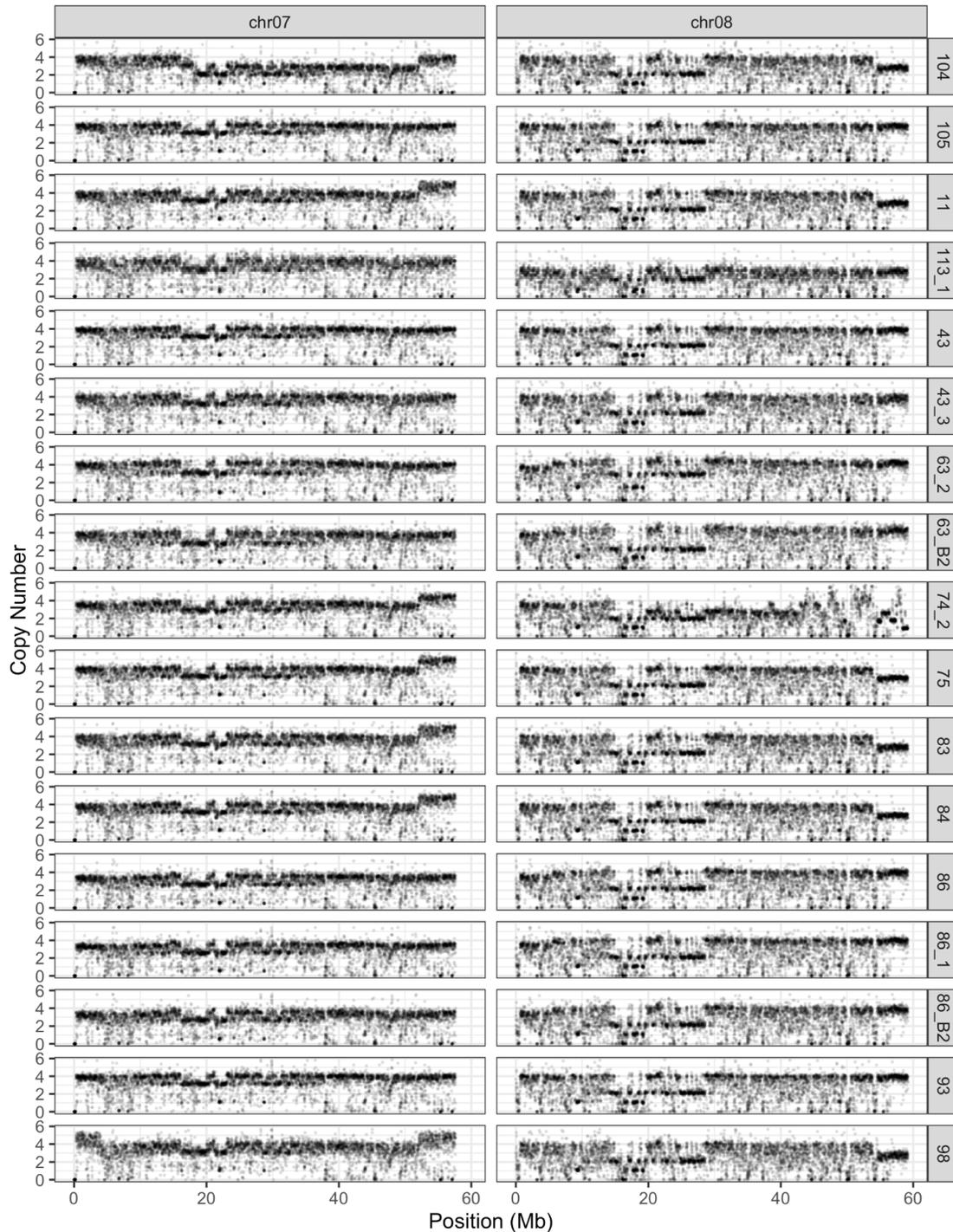
Supplemental Figure S4.2. Dihaploid linkage analysis. **A)** Expected segregation pattern and corresponding standardized coverage values among primary dihaploids of PI310467, given tr8-7. **B)** Overlay plots of standardized coverage values for 81 primary dihaploids of PI310467, 1Mb bin size. Traces from individual dihaploids are overlotted with high transparency. Dihaploids show recurring dosage variation of the long-arm ends of both chromosomes (shaded in green). **C)** Mean standardized coverage values for 81 dihaploids along the dosage variable termini of chromosomes 7 and 8, illustrating linkage between dosage haplotypes on the two chromosomes. **D)** tr8-7 junction identification via primary dihaploids with the Up7-Down8 or Down7-Up8 karyotype. Supplement to Figure 4.1.



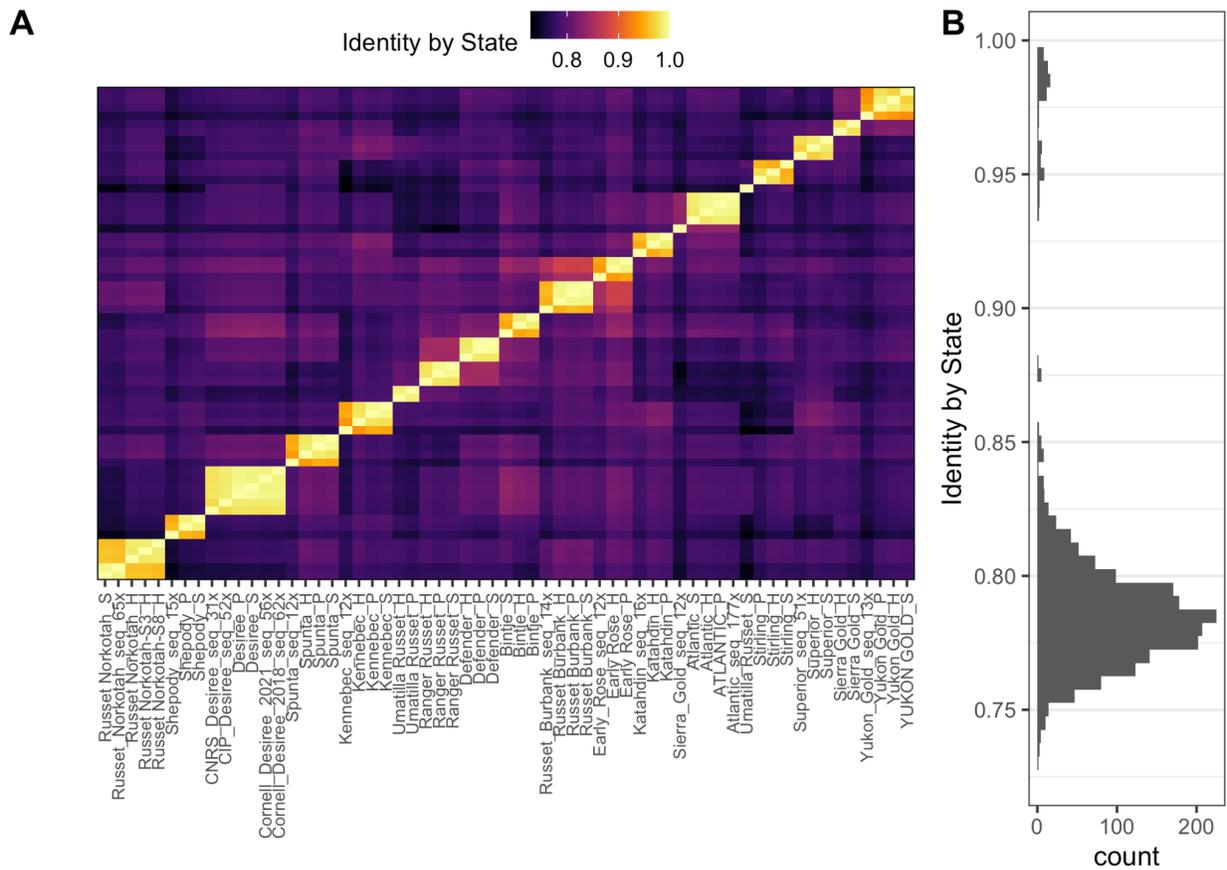
Supplemental Figure S4.3. PI310467 S1 population linkage analysis. **A)** Expected segregation pattern and corresponding standardized coverage values among PI310467 selfs, given tr8-7. **B)** Skim dosage plots of potato accession PI310467 tetraploid selfs. Traces from each S1 progeny are plotted with high transparency. For each S1 individual, dosage was measured relative to PI301467, which carries tr8-7. The copy number of up, mid and down states on chr07 correspond to 6, 5 and 4 copies, respectively. Similarly, the up, mid and down states on the end of chr08 correspond to 4, 3 and 2 copies, respectively. **C)** Swarm plots of average standardized coverage values for dosage variable termini of chr07 and chr08. Supplement to Figure 4.1.



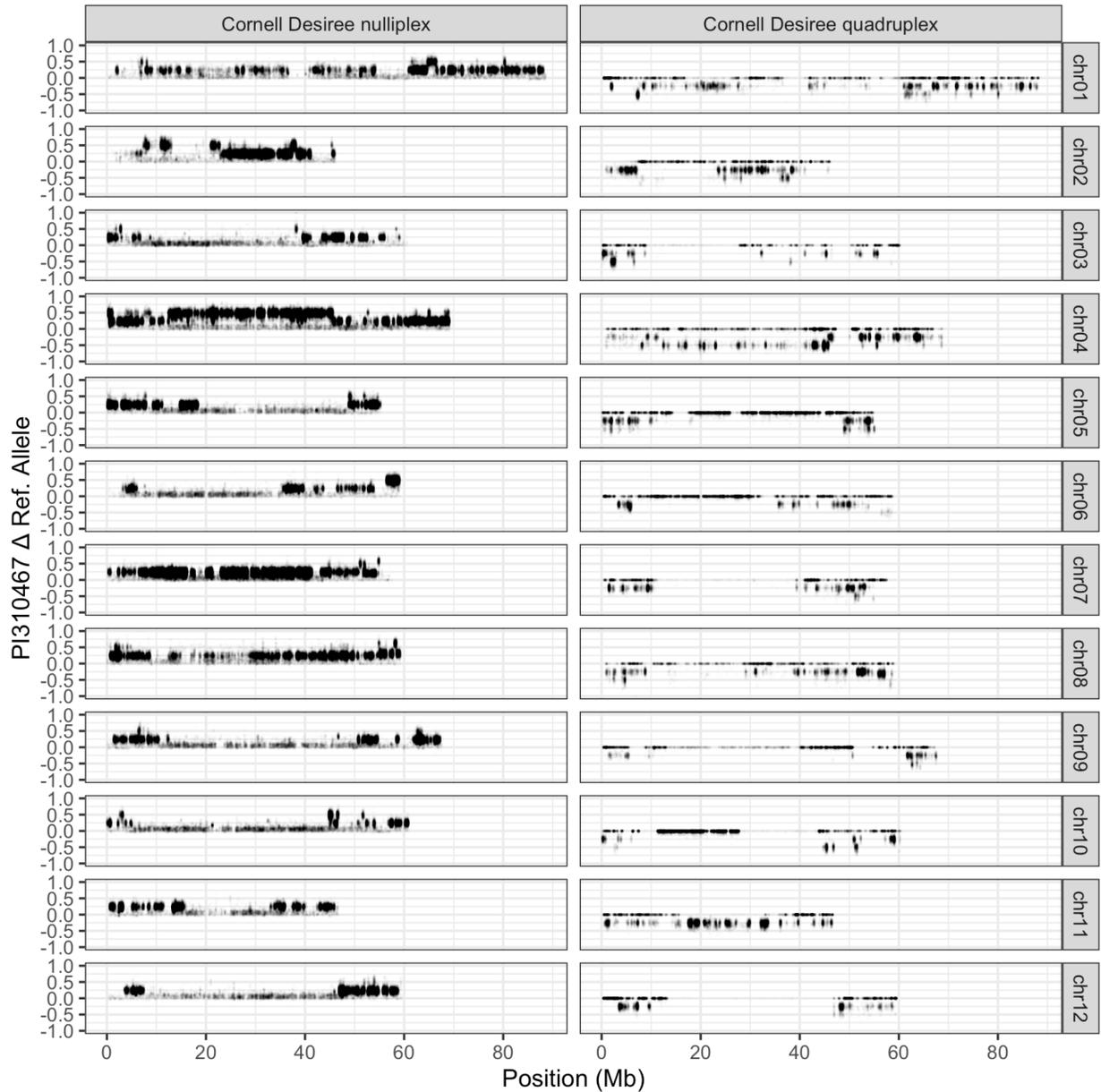
Supplemental Figure S4.4. Skim dosage plots of three independent cuttings from 2019 acquisition of potato accession PI310467. As read depth was standardized to PI310467, which has tr8-7, standardized coverage near 4 at the chr07 and chr08 ends indicates that tr8-7 is present. In contrast, tr8-7 reversion in regenerated clone MF93 appears as CNV relative to PI310467. Supplement to Figure 4.1.



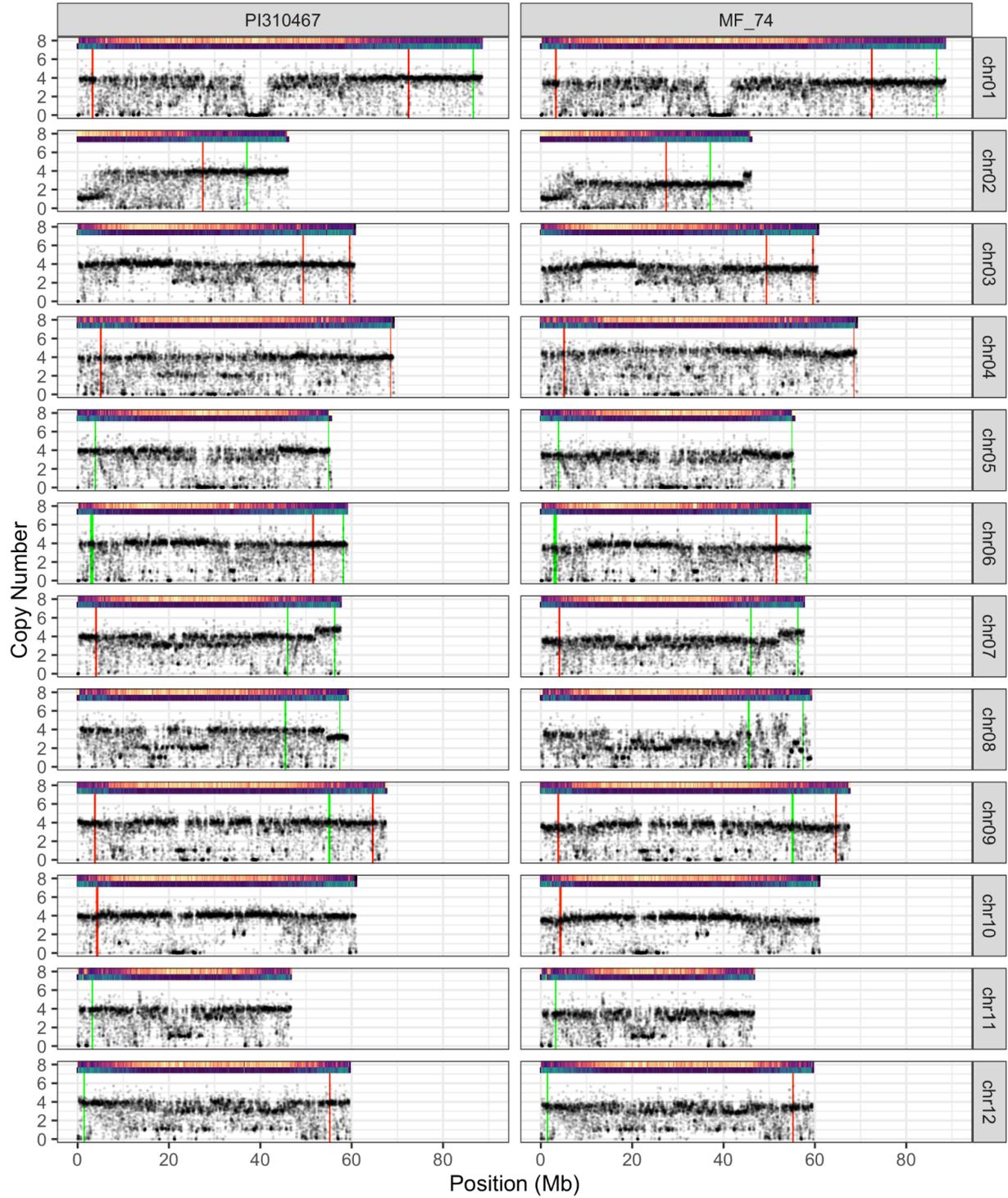
Supplemental Figure S4.5. Read depth analysis of regenerated clones. Chromosome 7 and 8 coverage plots of all protoplast-regenerated PI310467 clones. Median read depth values in non-overlapping 10kb bins are standardized internally to each sample and plotted at high transparency. Note the highly variable dosage pattern in the long arm of MF74_2. Supplement to Figure 2.



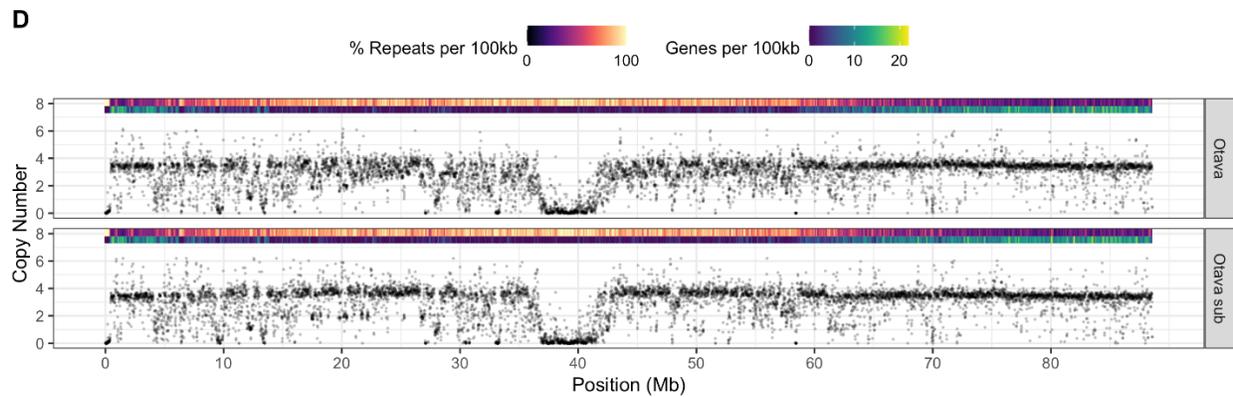
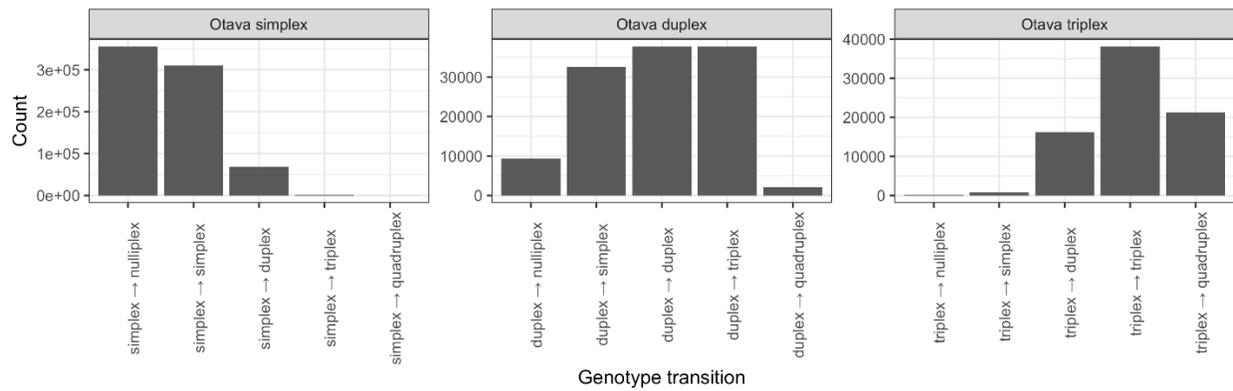
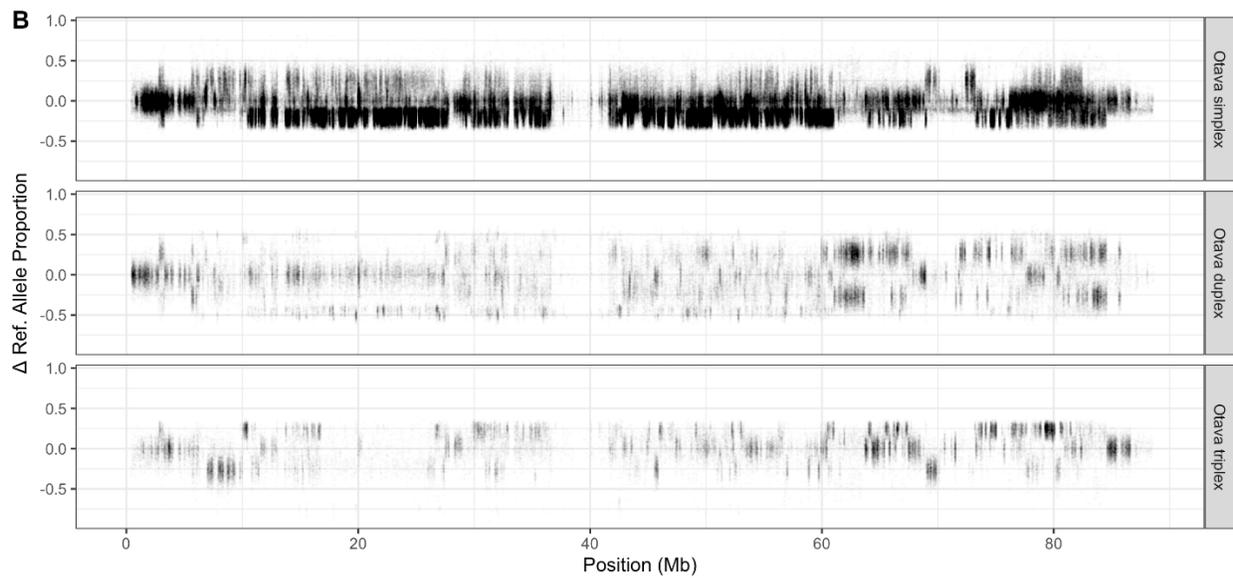
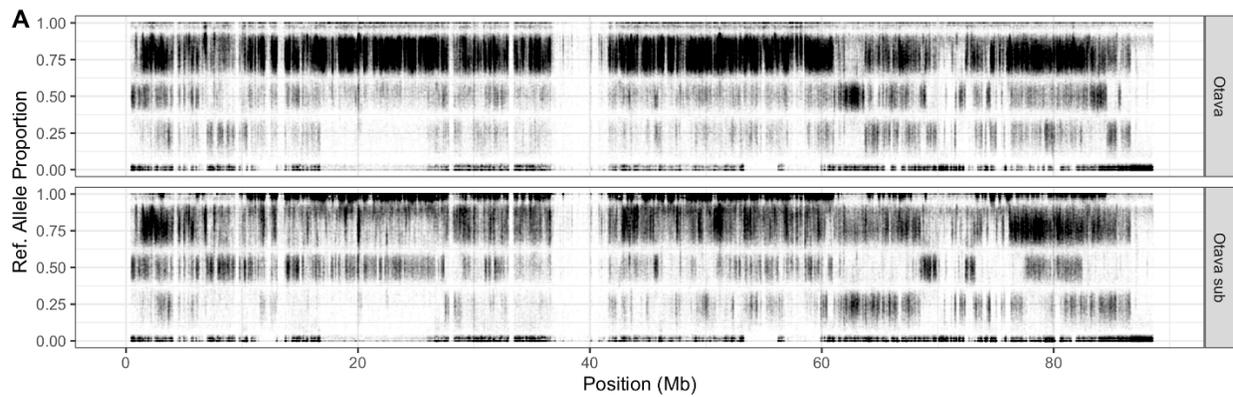
Supplemental Figure S4.6. Phylogenetic analyses by comparison of clones with multiple entries. A) Heatmap illustrating pairwise relatedness between clones with replicated entries in the panel of 1,020 potatoes, based on 1,858 SNPs. **B)** Histogram of IBS values from the matrix in panel A, bin size 0.005. Supplement to Figure 4.2.



Supplemental Figure S4.7. PI310467 heterozygosity at Desiree-homozygous SNP loci. The X-axis indicates reference genome coordinate, and the Y-axis is the change in reference allele abundance in PI310467, showing only loci that were homozygous in cv. Desiree from Cornell University and were biallelic in PI310467, all regenerants and cv. Desiree. Panels indicate maximum likelihood genotype of Cornell Desiree. Chromosome 8 is shown. Supplement to Figure 4.2.



Supplemental Figure S4.8. Expanded view of PI310467 and MF74 read depth. Coverage plot of PI310467 is shown on the left and protoplast regenerant MF74 is shown on the right. Gene and repeat densities based on DM1-3 v6.1 annotation are shown above each panel. Positions of FISH barcode probes from (Braz et al., 2018) are indicated with vertical lines on each panel, with the color of the line indicating probe color as previously described. Supplement to Figure 4.3.



Supplemental Figure S4.9: Copy-neutral change of heterozygosity from simulated haplotype substitution of cv. Otava. For both samples, only chr01 is shown. **A)** Reference allele dosage of Otava (upper panel) and simulated substitution (lower panel). Data points correspond to the proportion of reads supporting the reference allele at a locus and are plotted at high transparency. **B)** Change in reference allele dosage between Otava and simulated substitution line. Upper panels show dot plots displaying change in allele dosage with respect to reference genome position. Lower panels show histograms of genotype call concordance between Otava and simulated substitution line. **C)** Coverage plots of both Otava and the simulated substitution line. Data points correspond to median read depth of non-overlapping 10kb bins and are plotted at high transparency. Gene and repeat densities based on DM1-3 v6.1 annotation are shown above each panel. Supplement to Figure 4.4.

Supplemental Tables

Supplemental Table 4.1. Counts of PI310467 genotypes at 1,078,359 Desiree-homozygous loci.

Genotype class	Count
Desiree hom - PI310467 het	952,472
Desiree hom - PI310467 hom same allele	125,878
Desiree hom - PI310467 hom opposite allele	9

Supplemental Data Sets

Supplemental Data Set S4.1. Attributes of sequencing libraries produced or analyzed in this study

Supplemental Data Set S4.2. Identity by state matrix of 1,019 potato clones analyzed in this study.

Supplemental Data Set S4.3. Consolidated genotype calls from previous array genotyping studies and public whole genome sequencing.

Chapter 5

General Conclusions

Overview of Dissertation Research

The primary focus of this research was to investigate the effects of *in vivo* haploid induction and regeneration in tissue culture on plant genome stability. Using potato as a model system, the following questions were addressed: 1) Are the uniparental dihaploids obtained from haploid induction crosses of potato the result of double fertilization? 2) Among hybrids obtained from potato haploid induction crosses, what is the fate of chromosomes inherited from the haploid inducer parent? Do they show signs of instability? 3) What mechanisms account for changes in chromosome number and/or structure among plants regenerated in tissue culture?

Previous studies have documented additional chromosome(s) and haploid inducer-specific genetic markers in otherwise dihaploid potatoes. While these data are consistent with incomplete postzygotic chromosome elimination, an additional chromosome could also be due to female meiotic nondisjunction resulting in an aneuploid egg cell, meaning that the structure of any retained haploid inducer DNA. Across two studies, we identified and determined the parental origin of additional chromosomes and chromosome segments in 1,086 primary dihaploids by whole genome sequencing. As haploid induction is often incompletely penetrant, hybrids will also be obtained from haploid induction crosses. In centromere-mediated haploid induction of *Arabidopsis*, which is known to result from uniparental genome elimination (Ravi and Chan 2010), chromosomes inherited from the haploid inducer parent can exhibit simple or complex patterns of breakage and restructuring consistent with prior genome instability (Tan et al. 2015; Kuppu et al. 2015; Maheshwari et al. 2015). This research also addressed whether selective breakage of inducer chromosomes occurred in 134 hybrids produced by the incompletely penetrant haploid induction system of potato. Finally, regeneration of whole plants from single

cells is known to result in changes to chromosome structure and number, though the sequence of events underlying these changes has remained elusive. A panel of tetraploid potatoes regenerated from leaf protoplasts were characterized using genomic and cytological approaches to identify mechanisms that may explain widespread observations of genomic instability among regenerated plants.

Genome sequencing of a cohort of 167 primary dihaploids of Andigenum Group cultivar Alca Tarma revealed maternally inherited aneuploidy and widespread structural variation. Aneuploidy was detected in 11.2% of progeny, and was a single additional chromosome from the maternal parent rather than the haploid inducer in all cases. Meiotic nondisjunction offers a plausible explanation for maternally derived aneuploidy. The frequency and parental origin of aneuploidy in this population was consistent with previous studies (Wagenvoort and Lange, 1975; Samitsu and Hosaka, 2002; Pham et al., 2019). Addition or homology-dependent replacement of haploid inducer DNA segments has also been reported (Pham et al., 2019), but after imposing rigorous criteria for their identification, we did not find them. Consistent with previous reports of widespread structural variation in potato (Iovene et al., 2013; Hardigan et al., 2016; Pham et al., 2017; Hardigan et al., 2017), structural variation was common in the dihaploid population, and was generally due to segregating polymorphism inherent to the tetraploid seed parent. The largest of these were a copy-neutral rearrangement affecting the entire euchromatic short arm of chromosome 4, and CNV of the short arm of chromosome 2, the location of the nucleolar organizing region in potato (Dong et al., 2000). Rare structural variants were not attributable to the haploid inducer, and were explained by reference assembly errors, or by structural

differences between Alca Tarma and the reference genotype, DM1-3 (Potato Genome Sequencing Consortium et al., 2011).

A survey of a larger cohort of 919 primary dihaploids produced by pollination of 19 tetraploid genotypes with one of three haploid inducer genotypes again revealed aneuploidy at comparable frequencies to previous studies (Wagenvoort and Lange, 1975; Pham et al., 2019) that, in about 90% of cases, was again due to additional chromosomes from the tetraploid parent rather than the haploid inducer. Eight primary near-dihaploid aneuploids carried one or more chromosomes from the haploid inducer parent, indicating double fertilization had occurred in these cases. A search for addition or homology-dependent replacement of inducer DNA segments was conducted in three of the eight lines. While we could identify addition of inducer-derived haplotypes, it was not possible to determine whether these regions had integrated into the donor genome, remained autonomous or were associated with the trisomic, haploid inducer-derived chromosome. Selective instability of inducer-derived chromosomes in hybrids is a hallmark of CENH3-mediated postzygotic genome elimination (Tan et al., 2015). Genome sequencing of 30 triploid and 104 tetraploid hybrid progeny of haploid induction crosses revealed instability of haploid inducer-derived chromosomes, but this instability was ploidy dependent. All investigated haploid inducer genotypes showed ploidy-dependent genome instability. Specifically, whole-chromosome and segmental aneuploidy was rare in triploid hybrids, but frequent in tetraploid hybrids, suggesting mechanistic differences between CENH3-mediated and potato haploid induction. We considered three possible ways of producing a tetraploid hybrid: first meiotic division restitution (FDR), second meiotic division restitution (SDR), or restitution of pollen mitosis II (RS). By phasing the haploid inducer haplotypes in near-dihaploid aneuploids and

testing for heterozygosity of this haplotype in the tetraploid hybrids, we tested whether tetraploids were produced by FDR, SDR or RS. While approximately 95% of hybrids were produced by FDR, instability of haploid inducer chromosomes was not uniquely associated with any single mechanism.

Genome resequencing and cytogenetic analyses of 12 potatoes regenerated from leaf protoplasts was performed to investigate the mechanistic basis of chromosomal changes sustained during tissue culture regeneration. Outcomes consistent with chromoanagenesis and substitution of entire haplotypes were detected among the regenerants. A previous study of these clones identified recurring structural variants among regenerants from independent calli (Fossi et al., 2019). Population genomic analysis revealed the likely basis of this recurring variation: an unbalanced chromosome translocation that occurred in either a single cell of either the L2 or L3 layer of the protoplast donor, which ultimately became fixed in both cell layers. In contrast, the L1 retained an ancestral, balanced karyotype. Regeneration of L1- or L2/L3-derived leaf protoplasts into whole plants provides a plausible explanation for the recurring variation. These findings provided a snapshot of the genetic variation of long-lived, vegetatively propagated polyploids, due to mutations accumulated throughout vegetative propagation or induced by tissue culture.

The collective findings of these studies offer new insights into potential bottlenecks to potato improvement. In a population of primary dihaploids, less than 1% are expected to be near-dihaploid aneuploids due to incidentally retained haploid inducer DNA. Compared to the 8% of progeny expected to be maternal aneuploids, this frequency is low enough to be a minimal

concern for practical applications. Further study, especially on the cellular events associated with potato haploid induction may provide new insights on the underlying mechanisms, and potentially lead to development of a portable and highly efficient haploid induction system for eudicots, where such a system is currently lacking. Insights into genetic mosaicism and regeneration induced instability highlighted the impacts of prolonged vegetative propagation and regeneration-induced genome instability in potato. Understanding the broad extent of these effects will hopefully provide insights on methods for sustainable potato improvement with a combination of conventional breeding and emerging technologies.

Literature Cited

- Dong, F., Song, J., Naess, S.K., Helgeson, J.P., Gebhardt, C., and Jiang, J.** (2000). Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. *Theor. Appl. Genet.* **101**: 1001–1007.
- Fossi, M., Amundson, K.R., Kuppu, S., Britt, A.B., and Comai, L.** (2019). Regeneration of *Solanum tuberosum* plants from protoplasts induces widespread genome instability. *Plant Physiol.*
- Hardigan, M.A. et al.** (2016). Genome Reduction Uncovers a Large Dispensable Genome and Adaptive Role for Copy Number Variation in Asexually Propagated *Solanum tuberosum*. *Plant Cell* **28**: 388–405.
- Hardigan, M.A., Laimbeer, F.P.E., Newton, L., Crisovan, E., Hamilton, J.P., Vaillancourt, B., Wiegert-Rininger, K., Wood, J.C., Douches, D.S., Farré, E.M., Veilleux, R.E., and Buell, C.R.** (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. U. S. A.* **114**: E9999–E10008.
- Iovene, M., Zhang, T., Lou, Q., Buell, C.R., and Jiang, J.** (2013). Copy number variation in potato - an asexually propagated autotetraploid species. *Plant J.* **75**: 80–89.
- Kuppu, S., Tan, E.H., Nguyen, H., Rodgers, A., Comai, L., Chan, S.W.L., and Britt, A.B.** (2015). Point Mutations in Centromeric Histone Induce Post-zygotic Incompatibility and Uniparental Inheritance. *PLoS Genet.* **11**: e1005494.
- Maheshwari, S., Tan, E.H., West, A., Franklin, F.C.H., Comai, L., and Chan, S.W.L.** (2015). Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genet.* **11**: e1004970.

- Pham, G.M., Braz, G.T., Conway, M., Crisovan, E., Hamilton, J.P., Laimbeer, F.P.E., Manrique-Carpintero, N., Newton, L., Douches, D.S., Jiang, J., Veilleux, R.E., and Buell, C.R.** (2019). Genome-wide Inference of Somatic Translocation Events During Potato Dihaploid Production. *Plant Genome* **12**.
- Pham, G.M., Newton, L., Wiegert-Rininger, K., Vaillancourt, B., Douches, D.S., and Buell, C.R.** (2017). Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. *Plant J.*
- Potato Genome Sequencing Consortium et al.** (2011). Genome sequence and analysis of the tuber crop potato. *Nature* **475**: 189–195.
- Ravi, M. and Chan, S.W.L.** (2010). Haploid plants produced by centromere-mediated genome elimination. *Nature* **464**: 615–618.
- Samitsu, Y. and Hosaka, K.** (2002). Molecular marker analysis of 24- and 25-chromosome plants obtained from *Solanum tuberosum* L. subsp. *andigena* ($2n = 4x = 48$) pollinated with a *Solanum phureja* haploid inducer. *Genome* **45**: 577–583.
- Tan, E.H., Henry, I.M., Ravi, M., Bradnam, K.R., Mandakova, T., Marimuthu, M.P., Korf, I., Lysak, M.A., Comai, L., and Chan, S.W.** (2015). Catastrophic chromosomal restructuring during genome elimination in plants. *Elife* **4**.
- Wagenvoort, M. and Lange, W.** (1975). The production of aneudihaploids in *Solanum tuberosum* L. group *Tuberosum* (the common potato). *Euphytica* **24**: 731–741.